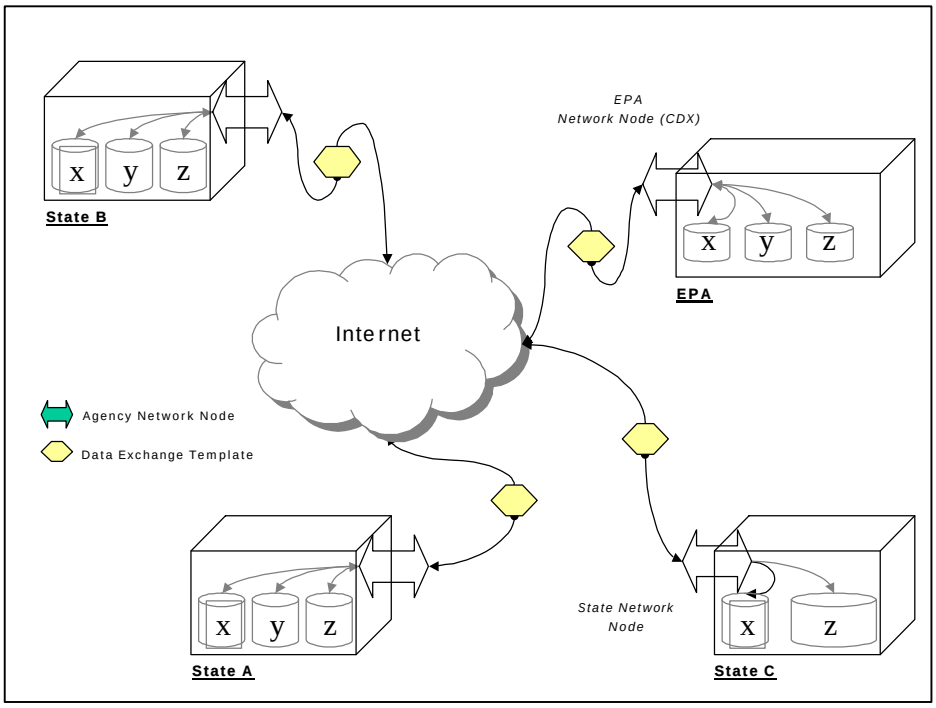
**Report to the State/EPA Information Management Workgroup**

# Blueprint for a National Environmental Information Exchange Network




## Prepared by the Network Blueprint Team
## October 30, 2000

## Document amended June 2001

# Foreword

The design of the Network is based on the observations and findings of the State/EPA Information Management Workgroup (IMWG), as outlined in *Shared Expectations of the State/EPA Information Management Workgroup for a National Environmental Information Exchange Network (the Network)*, June 2000 working version. (See Appendix B)

In July 2000, the IMWG formed a team charged with developing a Blueprint that would serve as the conceptual design of the Network. The intended audience for this document includes Chief Information Officers/Chief Technology Officers at state environmental agencies and their associated counterparts at EPA. The Blueprint Team was asked to 1) describe the Network's components; 2) test and refine the Network vision itself; 3) identify and assess the technical issues and options; 4) identify critical policy and political issues; 5) describe the specific, visible benefits we expect the Network to produce as it is implemented; and 6) finish as quickly as possible. This report to the State/EPA Information Management Workgroup documents the Blueprint Team's analysis and recommendations with regard to these issues.
The Workgroup received, deliberated on, and formally endorsed this report at its October 2000 meeting. The Workgroup authorized release of this final version of the report for wider distribution and further communication. In addition, the Workgroup charged the team to report back within six weeks of the final report with specific recommendations on the following:

1. Sourcing and/or establishment of the Network Administration function, including the possible use of third parties, funding, and its relationship to the Workgroup and the Data Standards Council.
2. Specific roles and responsibilities of the function of Network administrator and the means by which those roles and responsibilities could be fulfilled.
3. An implementation approach for these recommendations.

In February 2001, the Blueprint Team presented recommendations on the above Network Administration issues to the Workgroup. The Team's report was approved by the Workgroup and is included as an addendum to this document.

For answers to some Frequently Asked Questions about the Network, please refer to Appendix A: Network FAQs. A list of acronyms used in this document and a glossary of terms are included as Appendices C and D.

# Acknowledgements

This document has been developed through the support and analytic contributions of a number of individuals and programs within EPA, ECOS, and several state agencies. These individuals offered valuable insight, lessons learned from their experience in various agencies and organizations, and hard work in creating the Blueprint.

## State Participants

David Blocher, Maine Department of Environmental Protection
Dennis Burling, Nebraska Department of Environmental Quality
Troy W. Delung, Virginia Department of Environmental Quality
Ken Elliott, Utah Department of Environmental Quality
Irene Kropp, New Jersey Department of Environmental Protection
Renee Martinez, New Mexico Environment Department
Mark McDermid, Wisconsin Department of Natural Resources
Melanie Morris, Mississippi Department of Environmental Quality
Kimberly Nelson, Pennsylvania Department of Environmental Protection
Lynn Singleton, Washington State Department of Ecology
Ron Tuminski, New Jersey Department of Environmental Protection
Cathy Wagenfer, Maryland Department of the Environment
Mitch West, Oregon Department of Environmental Quality
Bob Zimmerman, Delaware Department of Natural Resources

## EPA Participants

Mark Badalamente, Office of Environmental Information
Andy Battin, Office of Water
Chris Clark, Office of Environmental Information
Connie Dwyer, Office of Environmental Information
Steve Goranson, Region 5/Office of Information Services
Lisa Jenkins, Office of Environmental Information
Chuck Spooner, Office of Environmental Information
John Sullivan, Office of Environmental Information

## Environmental Council of States

Mary Blakeslee

## Support Contractors

Matt Nielson, Concurrent Technologies Corporation
Andrea Reisser, Concurrent Technologies Corporation
Angela Jones, Ross & Associates Environmental Consulting, Ltd.
Lewis McCulloch, Ross & Associates Environmental Consulting, Ltd.
Louis Sweeny, Ross & Associates Environmental Consulting, Ltd.

# Table of Contents

# 1. Executive Summary

## A. Introduction

Information is fundamental to the work of environmental protection. State environmental agencies and U.S. EPA depend upon the rational flow of quality information for every aspect of their work, as individual agencies and collectively. Yet, many of the current systems and approaches to information exchange are ineffective and burdensome. This Network Blueprint describes a practical vision for an alternative to the current approach. It outlines a National Environmental Information Exchange Network (Network) that applies the technologies and approaches that have transformed the Internet to the exchange of data between environmental agencies. The specific technologies, and their application, are detailed in this Blueprint document. The core of the Network, however, is not technology: it is a commitment to change the way data is exchanged.

The Network will depend on the ability of environmental agencies to negotiate and then define the exact format in which data will be exchanged (data exchange template), to document the agreement in a trading partner agreement (TPA), and to hold parties responsible for fulfilling these agreements. Responsibility for data quality, timeliness, format and availability will be explicitly defined, documented, and agreed to by a designated individual for each party. Data originators will fulfill these agreements by maintaining information sources (nodes) on the Network that can provide this information upon authorized request. Once established, these data exchanges will replace (and be superior to) the traditional approach to information exchange that relied upon states "feeding" information directly to EPA's national data systems. Those agencies that choose to utilize the Network would do so in place of their traditional “feed the system” uses of national systems at EPA.

## B. Background

The analysis and discussion reflected in this Blueprint involved a team of more than 40 State and EPA staff, as well as associated contractors and technical experts. Given the complexity and diversity of existing flows, this transition will be gradual, but accelerating. New and old approaches will necessarily exist side by side for many years. Guidance for managing these transitions will emerge only through actual experience. The recommendations at the end of this document constitute a proposal from the Network Blueprint Team to the IMWG to begin this joint effort now.

## C. Challenges and Opportunities

A joint commitment to implement this Network clearly carries challenges and risks: these are described in the document. Inaction also carries risks. Regardless of this Network, states, EPA and other potential partners are making, and will continue to make, investments in new systems designed to fit their business needs. In most cases, this will mean that EPA national systems will no longer be primary operational systems for states (and others). Without a compelling and

credible organizing framework for how to share information in this new world, the quality and reliability of those collective efforts will be at risk and a unique opportunity for joint progress will have been missed.

Within this Network Blueprint document, the key remaining issues to be resolved cluster around the administration of the Network itself and the logistics of converting historical system-specific flows to Network flows.

## D. Network Design

The Network is based upon four basic principles. These principles were developed in the Shared Expectations document and have remained intact:

- Stewardship of specific data will be established by mutual agreement between two or more trading partners.
- Stewards, through their node, are directly responsible for the quality and availability of this data.
- Network members whose use of stewarded data necessitates the maintenance of local copies are responsible and accountable for ensuring the integrity and currency of those copies.
- Network members agree to use the Network technology standards, as described (and refined) in this Blueprint and as documented in their individual trading partner agreements. These principles are implemented through five components: 1) Data Standards, 2) Data Exchange Templates, 3) Trading Partner Agreements, 4) Technical Infrastructure and Network Administration, and 5) Member Organizational Infrastructure.

## E. Recommendations

The Blueprint Team ultimately envisions a broad and diverse membership, including local, state, federal and tribal agencies. The Blueprint Team also envisions the Network beginning with states and EPA and expanding as fast as experience and the interest of others allow. This Network is expected to dramatically improve the quality and availability of environmental data to environmental agencies and the public. The Blueprint Team recommends that the IMWG formally and fully endorse this Blueprint. Further, the IMWG should charge the Blueprint Team with developing and forwarding a specific proposal for how the Network administrative function, including financing options, should be established. Finally, the IMWG should identify its next steps in advancing the Network, including a plan for outreach, and recognize that these steps should begin immediately.

Note: this Blueprint Report was endorsed by the IMWG at its October 2000 meeting. Per this recommendation, the Workgroup charged the Team with developing a specific recommendation for how the Network administration function should be established. The team delivered this final product, included here as an addendum, in February 2000. The IMWG endorsed all recommendations in the proposal.

# 2. Introduction

Information is fundamental to the work of environmental protection. State environmental agencies and U.S. EPA depend upon the rational flow of quality information for every aspect of their work. Yet, many of their current systems and approaches to information exchange are ineffective and are overly burdensome, with obsolete and expensive computer systems that satisfy neither staff nor external users (e.g., the public, regulated industries). At the same time, two significant trends exacerbate the need for a new approach to environmental information systems. First, environmental protection agencies collect, access and utilize increasingly more environmental data, as the scale and complexity of the problems addressed has grown. Second, a widening system of environmental information exchanges has already evolved with the devolution of management from the federal to the state and local levels.

In response to these trends, and to the growing expectation that this information and government services themselves will be available online, EPA, states, and others are making major new investments in information systems. The pace and intensity of these changes have brought the problems with the traditional system-to-system approach into clear view. As states and EPA make these new investment decisions, they have asked for a framework that can coordinate their efforts and build on a common vision. This Blueprint is intended to provide this framework. Specifically, state environmental agencies and EPA have struggled with modernizing systems at different paces, making it difficult to maintain the traditional direct system-to-system exchanges.

The rapid growth of the Internet and electronic-commerce (e-commerce) now provides a solution–an Internet-based voluntary National Environmental Information Exchange Network (Network) for state, federal and tribal environmental agencies. A Network based on standardized Internet protocols allows individual agencies to invest in internal data storage systems of their choice at a pace they can afford, while also supporting easy exchange of environmental data. Although the drivers and capability to create such a Network are already in place, its development will require *deliberate and collaborative* design and work. These areas are the focus of this document.

In overview, the Network facilitates information exchanges between “nodes” maintained individually by participating partners (initially envisioned as state environmental agencies and EPA). These nodes use the Internet to exchange information via standardized data exchange templates (DETs), using common (Internet-based) protocols. Exchange of data is governed by trading partner agreements (TPAs) between the partners. TPAs document the agreed upon data, exchange format, frequency of exchange and related issues. For example, a state and its EPA Region negotiate a Performance Partnership Agreement (PPA) that includes a TPA for the exchange of permitting, enforcement and compliance data for the National Pollutant Discharge Elimination System (NPDES) program. This TPA explicitly defines the quality, timeliness and format of the data, binding the state and EPA Region in a "data-centered" agreement. Held together by such agreements, the Network will bring clear and measurable benefits:

- A common approach to environmental information exchange that is manageable by an agency as an agency, and not a collection of stovepiped systems, loyalties and approaches.

- A transition from traditional information exchange approaches, which are rife with management and data quality problems, to a data-centric approach focused on data and data quality.
- Enhanced potential for data integration.
- Lower cost to exchange data.
- More agency control over its own data, especially in light of public and legislative trends driving all public data onto the Internet.

The approach and benefits envisioned for the Network have already been validated in the private sector, such as RosettaNet (see reference document on RosettaNet).

The Network approach also explicitly recognizes the ownership and responsibility of agencies for their data, and the responsibilities of participants who aggregate that data. By moving pro-actively to create this Network, participants can establish their nodes as the sources of record rather than have piecemeal or prescriptive approaches legislated or otherwise mandated. Although not a panacea for all existing problems, the Network allows more focus on interpretation of the data and, in turn, enables better environmental decision-making.

Initially, the scope of the Network will be limited to information that partners are already exchanging on a formal basis (e.g., states with EPA); vastly more agency data may be available on public access websites, state clearinghouses, and other informal arrangements than on this Network. As indicated above, flows of environmental information involve an ever-increasing number of governmental agencies (local and national). While this Blueprint focuses on state, EPA and tribal information flows as a starting point, it will expand to other participants as their interest and the capacity of the Network allow. The ultimate vision is a broad and diverse web of quality information, but the design begins small.

Figures 1 and 2 compare the more complex and disjointed process of data flow typical today with a more streamlined and efficient process that would occur on the Network. The most important aspect to note about these figures is the shift from the use of many transfer mechanisms between the states and EPA today to a much more standardized mechanism envisioned on the Network. Beyond improved data quality, consistency and coverage, this change will allow all Network participants to achieve economies of scale as they consolidate the function of information exchange and standardize the format of data to exchange.

**Figure 1: Overview of the current information reporting relationship between states and EPA.**

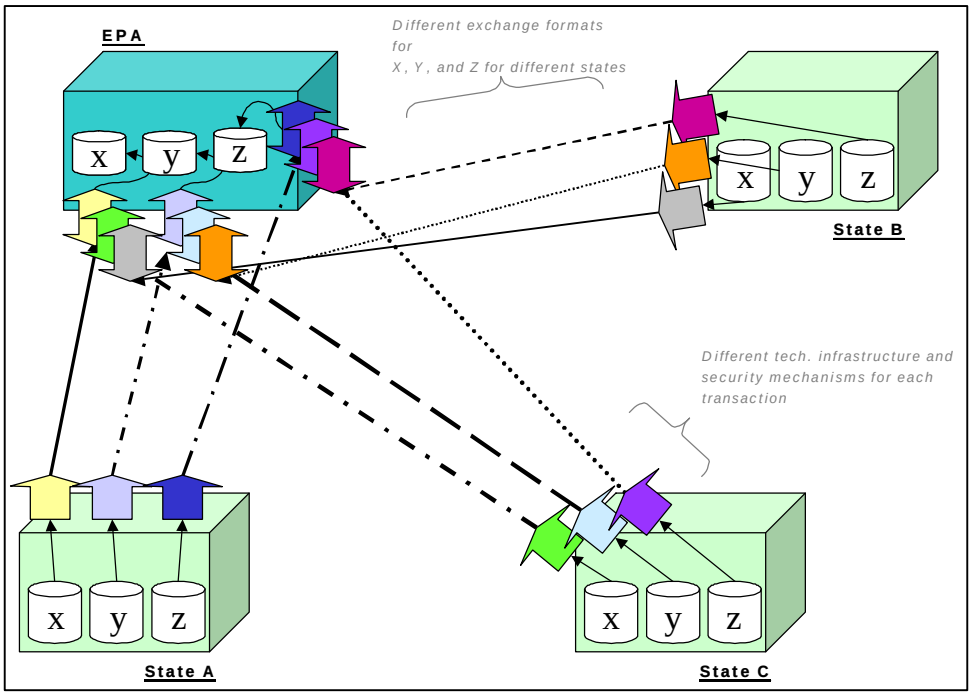
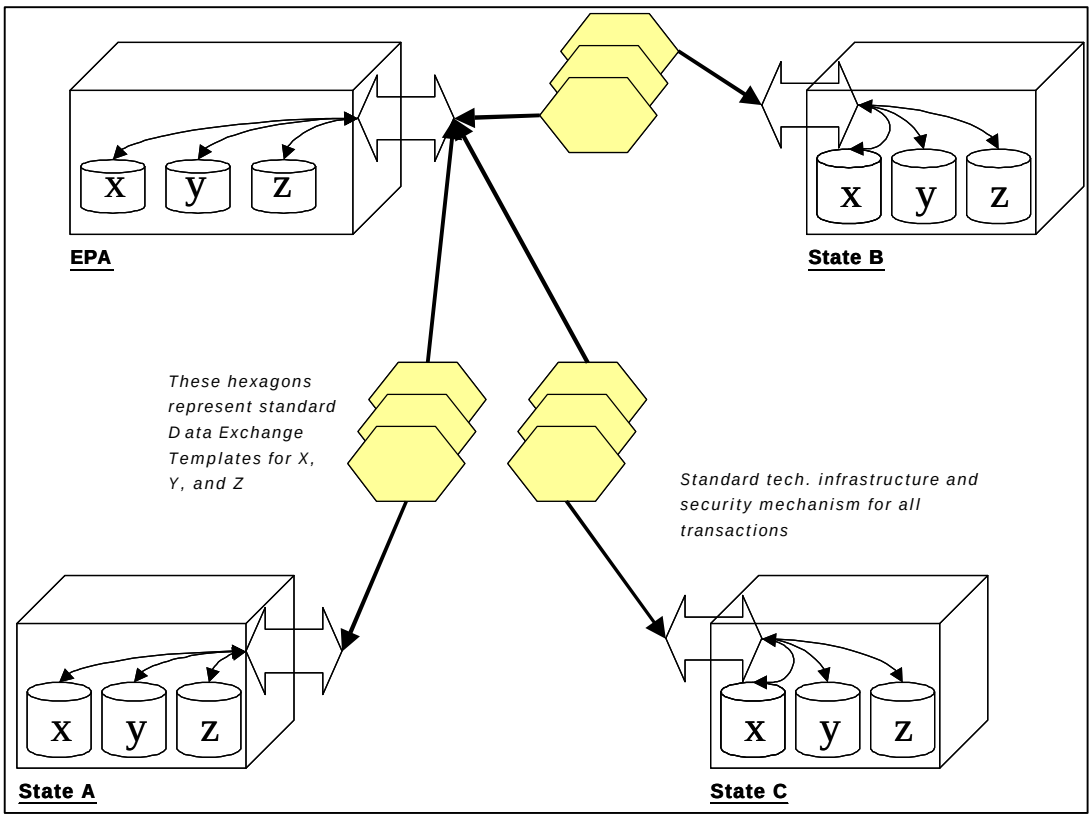



**Figure 2: Overview of the envisioned information reporting relationship envision between states and EPA**



The design of the Network is based on the following observations and findings of the State/EPA Information Management Workgroup (IMWG), as outlined in *Shared Expectations of the State/EPA Information Management Workgroup for a National Environmental Information Exchange Network*, June 2000 working version:

- Information, especially integrated information, is an increasingly important environmental management tool.
- Currently, this information is widely dispersed across state and EPA departments and locations and yet is increasingly demanded by a wide and diverse audience in an integrated fashion.
- Many states are investing in their own information systems and migrating away from use of EPA national systems.
- EPA faces the challenge of an increasing diversity of state and other data partner systems, ranging from those who have built integrated modern systems to those who continue to rely on EPA-sponsored systems.
- The current discussion concerning data among states and EPA is nowhere near as productive as it could be. The current collective approach leaves much to be desired in establishing clear accountability and responsibility for data quality, stewardship and management on all sides. Often these debates fail to even escape from disagreements over the definition of basic terms, or the currency or authority of given data sets or reports.
- There has been a revolutionary convergence of technologies around the Internet, World Wide Web (WWW) and e-commerce, especially the establishment of secured networks of standards-based information flows, which use the Internet as its infrastructure.
- Governments can apply these technologies to data they exchange with their partners, but governmental and intergovernmental coordination presents unique challenges to their use.
- A Network Blueprint is needed to allow shared and clearly defined terminology in addition to accountability and responsibility for elements such as data quality, timeliness and authority, exchange formats and methods, and access. This will allow each partner to operate independently on internal matters and in a coordinated fashion on external issues.

## A. Why a Network Blueprint?

The Shared Expectations document raised both significant interest and questions among IMWG members and their staff. "How would this work?" "Who would do this seriously, what must we start doing now?" In response, the IMWG commissioned an ad-hoc team of state and EPA staff to develop a conceptual Network design, the need for which became especially acute as the IMWG itself, EPA, and individual states began incorporating these concepts into their own investment and management decisions.

Ultimately, the Network will be whatever those who build and use it create. The pace of its evolution will be uneven among users.

The Blueprint is designed to support two essential next steps, without which the Network will not evolve (at least not from this effort):

1) A vigorous dialogue on the merits of and approaches to growing such a Network among states and EPA and tribes (to start).
2) Immediate support for those who will start building the Network. These efforts will start small, beginning with single data flows between two parties.

Within the context of the IMWG, this Blueprint is designed to support dialog and implementation at several levels:

- EPA, as it continues to refine its information strategy and near-term investments. If EPA accepts these Network concepts, investments in them will form a core strategic principle of its information strategy.
- Individual states, as they accelerate investments in information interchange, portals and e-commerce.
- The State EPA Information Management Workgroup, which seeks to coordinate state and EPA efforts.

The level of detail in this document varies widely from section to section, providing only enough detail to establish the plausibility and desirability of the Network parts. Substantial revision is expected before the design can be considered complete. Furthermore, the programmatically challenging aspects of the Network (e.g., the details of trading partner agreements) will require on-the-ground experience before refinement is possible.

## B. Overview of the Organization of the Network Blueprint Document

At its simplest, any network is made up of nodes and relationships (data flows and agreements) between those nodes. All the elements of the network – its infrastructure, policies and technologies – can be related back to these two fundamental parts. The Network Blueprint is organized as follows:

Section 3 provides a very high level overview of the Network concept and its parts.

Section 4 defines a Network node and describes its operation.

Section 5 describes stewardship of the data and the Network.

Sections 6-10 describe the components of a Network flow.

- 6: Data Standards
- 7: Data Exchange Templates
- 8: Trading Partner Agreements

- 9: Technical Infrastructure
- 10: Organizational Infrastructure.

Each section follows a common organization:

- A. Background: Basic context for the component.
- B. Definition: A brief definition.
- C. Business Case and Critical Features: The rationale for that component (i.e., why it is needed and what it does).
- D. Government Issues: Specific governmental issues raised by the component, especially where a private sector concept is being adapted to a government context.

Section 11 describes how the various Network components relate to each other.

Section 12 presents the recommendations being forwarded in this Blueprint.

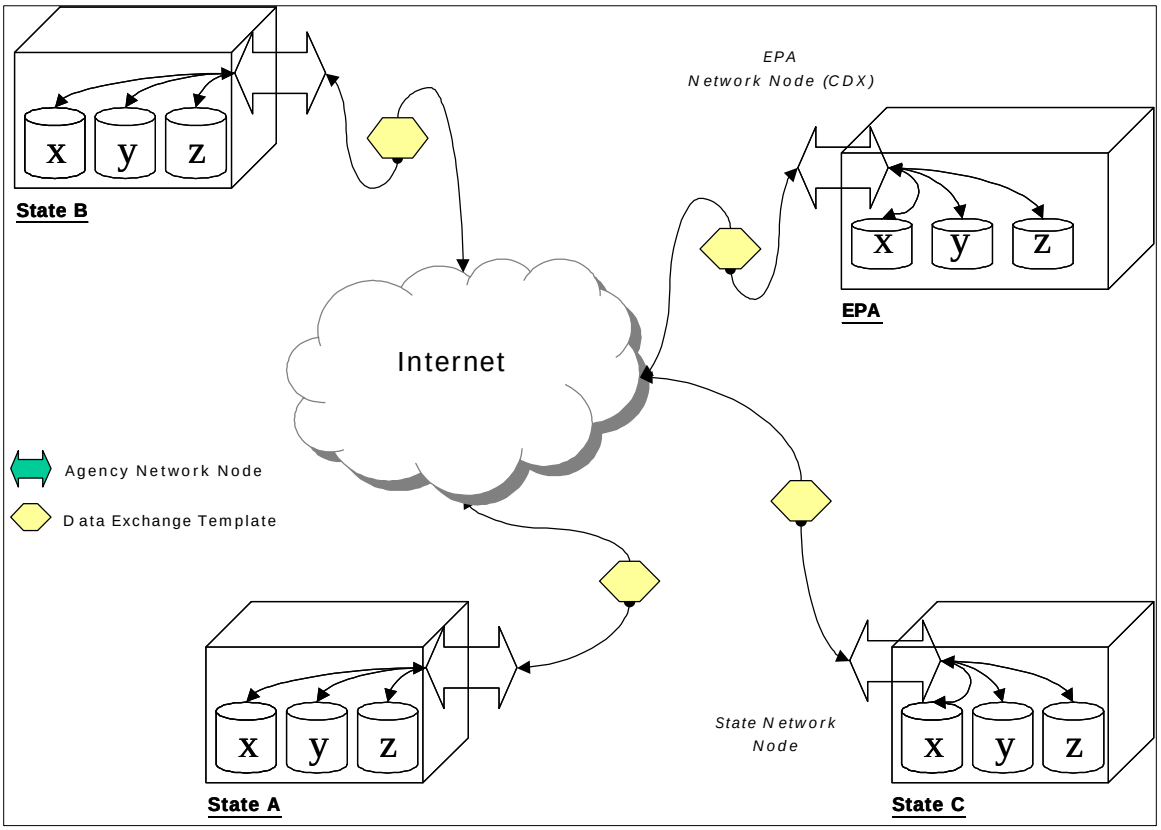
Section 13 presents a Network example.

Section 14 is the full text of the Recommendations and Initial Implementation Proposal on Network Administration delivered to the IMWG on February 6, 2000.

# 3. Overview of Network Design and Design Principles

In overview, the Network facilitates information exchanges between nodes maintained individually by participating partners initially envisioned as state environmental agencies and EPA. As shown in Figure 3, these nodes use the Internet to exchange information via standardized data exchange templates (DETs), using common (Internet-based) protocols. Exchange of this data is governed by trading partner agreements (TPAs), not shown, which document the agreed upon data, exchange format, frequency and related issues.

**Figure 3: Conceptual diagram of the Exchange Network**




The proposed Network design balances three sometimes-conflicting requirements:

- Desire for an open, dynamic, diverse network of environmental data flows, with an absolute minimum of constraints and overhead on participation.
- Reliable flows of data that is consistent nationally (and/or at other scales) that can be readily accessed and integrated.
- Capacity to fulfill the majority of participants' exchange obligations, whether regulatory, statutory, grant or otherwise.

The Network proposed in this Blueprint involves design requirements and compromises similar to those of the Worldwide Web (WWW), which is diverse and easily joined. Yet underlying it (and mostly invisible) is a strict, technically rigid set of standards for the transmission (TCP/IP)

and display (HTML) of information. There is flexibility in some places, but absolutely none in others. For example, the rules of the WWW preclude a nonconforming Internet address (e.g. 207.193.**green**.99.47), or a page that uses a proprietary variant of HTML. It just will not work. These technical design restrictions dramatically constrain what the Web can do, yet are wildly successful.

The proposed Network here manages the conflicts identified above by using the technical infrastructure of the Web to move standardized sets of information in agreed-upon DETs, and where necessary, to officially document the agreement to do so in a TPA. Some DETs will be created and used by only a small number (maybe just two) Network members. Other DETs will likely be adopted by all members using a Network flow to satisfy their obligations to a single member (e.g., states to EPA reporting under a delegation agreement). For EPA's traditional reporting flows, these DETs would function as national standards, but they would be only one part of a diverse and constantly expanding set of standards. They would be superior to the current approach because they would be expressed in uniform, unambiguous and self-validating formats, rather than through a process of "feeding" a legacy system.

# 4. What is a Network Node?

## A. Defining a Network Node[1]

A Network node is a participant's single, managed point of interaction between trading partners on the Network. The node is the collection of specific technical and policy components that a participating member will manage for providing and receiving information via the Network. Nodes have the following critical features:

- Each Network member has only one node, although that node may handle many kinds and types of data.
- A member's node is the only route for Network delivery and official receipt of information.[2]
- The node is the single place for each member to present its standard node catalog of available information and associated Network metadata (e.g., their TPAs, description of the information). Data and associated information must be presented on a node to be on the Network.
- The node is the single place where each member implements the minimal but essential transport, security and query protocols described in the Blueprint and specified in the TPA.
- The node is the only place where a member's compliance with a TPA can be demonstrated or evaluated.

Members may choose to link their nodes with their public access websites, but nodes and websites perform different functions and Network members will be required to ensure that adequate security is in place to separate the functions. Placing quality information on an attractive, well-designed public access website is a good thing – most agencies are doing this – but a website is not a node. A node presents this information, expressed in an extensible markup language (XML), using a standard DET and governed by a TPA. Figure 4 illustrates the functional differences between an agency's standard website and specific Network nodes. Unlike a standard website, flow through the node involves a specific request from a particular trading partner (not anonymous) for information listed or referenced on the node catalog, governed in a TPA, and presented in the correct format specified in a DET.

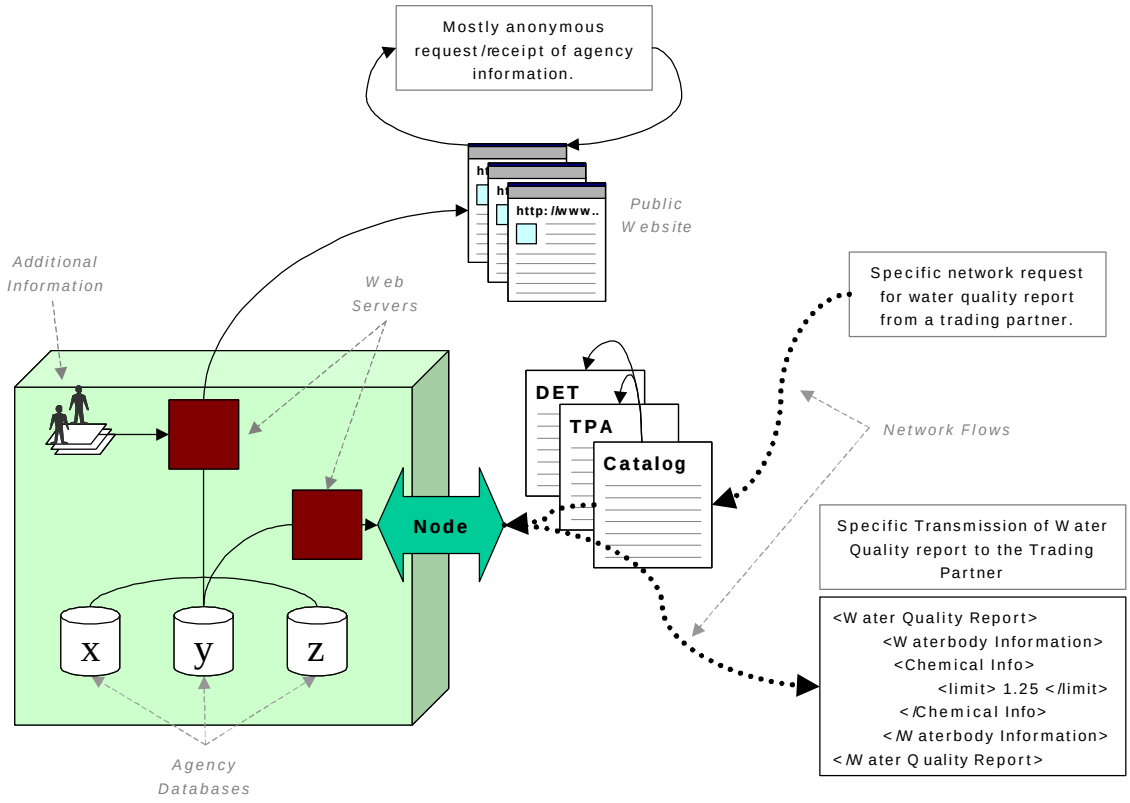
### What Is a Node "Really"?

A node is the central management point for each agency's interaction with the Network. All current flows take a program office-specific, system-specific, state-Region-specific path. This flow is difficult to manage, and the Network concept assumes the following simplification:

[1] The term "node" may also be known as "web services."
[2] Members may choose to make some information on their nodes publicly available and/or to use their nodes as "back-ends" for public access websites.

- All Network data (e.g., submittal of a quarterly report) flows from the originator's node
- These flows are governed by a TPA signed by a single authorized individual from each partner.

**Figure 4: Conceptual diagram of the internal operation of a Network Node.**




The transition to this new approach will not be easy or quick. Existing flows of information and Network flows will co-exist for several years. Managing a state or EPA node will require a new attention to internal roles and responsibilities that have much history. These roles are discussed further in Section 10: Member Organizational Infrastructure. They are introduced here because they touch on nearly every other aspect of the Network.

Today these flows are not from "state" to "EPA" as trading partners, but rather from the state's program, through a separate state system (or double entered), into to a specific program office and system within EPA, with the involvement of a specific person in the EPA Regional Office. This entire flow may be covered by several overlapping PPA, SEA (State/EPA Agreement) and program delegation agreements that are signed by various state and EPA officials to feed program data into an EPA system. These arrangements often vary greatly from state to state, Region to Region and program to program. The locus of authority and responsibility for data quality in this system is unclear. The Network, specifically the TPA, will not attempt to ignore the complexity of current flows; instead, it will simply make explicit the operation and obligations of nodes for a specific flow.

For all Network flows, the point of accountability for performance will be shifted from the program-specific data flow (e.g. manual entry into a federal system) to the state node. This clarity is an essential feature of the Network. Trading partners will no longer be left arguing about issues caused by ad-hoc movements (or manual double entry) of data into the other's system, but can instead focus on the quality and availability of data as specified in the TPA. It is expected that these arguments will be displaced by discussions about what the data say is happening in the environment.

## B. Node Operation, State Nodes and Central Data Exchange (EPA's Node)

Network members will build their nodes as an extension of their existing Web and enterprise architectures. As outlined above, the node has a set of relatively simple technical functions, but its key role is as a management point for data. This role is likely to require additional or new roles and relationships for EPA and state staff. EPA and many states have already begun investments in Web portals that draw information from their official enterprise production systems for public availability. This is very similar to the Network node except that the data would be bound under a TPA and formatted according to the DETs. In addition, each participant would be required to have a formal process for managing the flow of data from the production systems to the portal, since those flows would be official. What was once a person-based flow from one program office in the state to one system at EPA becomes an enterprise flow, both within the originating state and at EPA as data flows through EPA's Central Data Exchange (CDX). See the discussion of roles and responsibilities in Section 10 for more information.

In addition to servicing authorized requests for specific data, each node must be able to provide its catalog to authorized requestors. There are many approaches to formatting and providing the node catalog metadata. As in the case of the Trading Partner Agreement Markup Language (TPAmL) used as the basis for the TPA section of this document, many open source approaches (e.g., the Federal Geographic Data Committee [FGDC] "node" reference format) can be adapted easily to the Network.[3] This "cover sheet" or "lobby" to a node would allow participants to determine what data was available and how, if they are authorized, to access it. At its absolute simplest, this catalog could simply be a single XML file that is always found at the root level of a Network node's URL with a common agreed-upon name.

### How Will Nodes be Built and Operated?

The Network design is patterned after demonstrated approaches taken in the private and mixed sectors (e.g., healthcare). Some of the base technologies are young (e.g., XML), but as the Blueprint Team's analysis and independent expert consultations suggest, there is enough experience to support their use. These technologies are described in more detail in Section 9. In overview, participants will build their nodes as an extension of their current enterprise systems. Because the Network will be based on open standards (i.e., not tied to a particular technology or vendor) like XML, participants will be able to build their nodes using a wide and rapidly developing choice of tools. All major software vendors have now embraced these technologies,

[3] As anticipated in 2000, the ebXML (and other) initiatives have converged two standards for this information – WSDL (Web Services Description Language) and UDDI (Universal Data Discovery and Integration).

and many new companies have introduced products that make this market highly competitive and diverse. Perhaps most important, participants are free to implement any tool and any internal architecture for their node – the standards of node function are based purely on performance.

# 5. Stewardship

The flow of quality data is fundamental to the Network. The concept of stewardship refers to the responsibility for this data quality on the Network. As discussed above, this document seeks to resolve the current ambiguity in many data flows by establishing DETs and TPAs. Effective stewardship is essential for the Network to be successful, which will be achieved as Network data becomes synonymous with "high quality" data. Members will take responsibility for the data they place on the Network and for their interactions with the Network itself.

The concept of stewardship is involved in all of the principles and components of the Network. This section emphasizes some of the most important of these forms of stewardship.

## Data Stewardship

By agreeing to host and exchange data and information, each trading partner on the Network assumes and accepts certain data stewardship responsibilities:

- Assuring that responsibilities for data quality and integrity are clearly defined and understood inside the organization.
- Assuring that data source, derivation, and accuracy meet specifications.
- Assuring that data formats and units of measure meet specifications.
- Assuring that any other relevant data or metadata meet the specification in the TPA.

## Node Stewardship

Each partner, whether state, tribal or federal, will be the steward for its own node, making sure it functions properly and that the data available complies with agreed-upon terms.

Network Node

- Assuring that the hardware and software that create, manage, store and provide access to the data work properly.

Transmission/Transaction

- Assuring that the data transmitted and received are complete.
- Assuring that the data transmitted and received comply with agreed-upon formats and time schedules.
- Assuring that data have not been altered.
- Assuring that confidential and sensitive data have not been intercepted.

TPAs will ensure that data quality requirements are built into each data exchange, including quality, format standards, documentation standards, content, sources, accuracy and timeliness, error detection and correction methods, other conditions that affect acceptability of the data, and reconciliation of data quality concerns. The technical infrastructure component detailed in Section 9 describes how the technology supports this stewardship.

Some participants who make data available on the Network will not be the original authors of the data. In these cases, the role would be custodial – to store data for the convenience of access and analysis, with no attempt to govern the data or improve its quality.

## Stewardship of Registry Data

Registries – reliable and authoritative sources for commonly used data or code sets made available on the Network – will require shared stewardship across the relevant members. Because of these coordination needs, registries will present special stewardship challenges. One of the first registries to be established on the Network may be the regulated facility registry (FRS), maintained by EPA. Over time, EPA and perhaps other Network participants will expand existing registries and add new registries.

## Stewardship of Data Not on a Member's Node

The basic Network concept assumes that each trading partner can manage its own data and make this data accessible via its own node on the Network. This capacity will evolve incrementally from state and EPA investments. In some cases, member capacity to steward their data may mature before they have a node operational. For example, EPA's systems are used as the official systems of record for some states, including those with delegated programs. If EPA establishes the technical infrastructure for its node and is technically able to place this data on a "hosted" node for that state (for the state's, EPA's and other members' benefit), that state might choose to execute its stewardship through that national system. In this case, states would take on data stewardship and node stewardship would be shared.

# 6. Component 1: Data Standards

## A. Background

Data standards support the efficient and accurate exchange of data and assist secondary users to understand, interpret and use data appropriately. Note that these standards will apply to the "data" itself and to the "metadata", which provide additional information above the data/data set.

States, EPA and tribes recently established the National Environmental Data Standards Council to promote the identification, development and adoption of data standards. The Network will promote and acknowledge the use of all available standards developed or endorsed by the Data Standards Council. No other mechanism for creating or recognizing data standards is envisioned. The Environmental Data Standards Council has prioritized the standards that need to be developed and chartered workgroups (made up of additional state, EPA, and tribal members) to begin this work. Final standards will be posted on a website, available to all environmental agencies and trading partners. Most importantly, these standards will be used by participants to build DETs.

## B. Definition

As defined by the Data Standards Council, data standards are "documented agreements on formats and definitions of common data." Data standards are established to bring better consistency and quality to the information that organizations maintain.

Data standards provide the definitions and formats of the individual data elements (or "word Data elements alone are usually meaningful only when placed in data groups (or "sentences"). For example, the data element "mailing address line 1" is grouped with several other data elements, such as city name, state and zip code, to create the data group "mailing address." Some data standards also provide information about the interrelationships of its data groups.

The traditional components of a data standard are defined below.

- Data element – one particular piece of data; for each data element the following information is traditionally provided.
  - Name (e.g., Mailing Address Line 1)
  - Format (e.g., string, integer, date)
  - Definition
- Data group – logical grouping of data elements (e.g., the "Individual" data group in the Facility Identification Data Standard is made up of the elements "First Name, Last Name, Middle Initial, and Title Text")
- Relationships – the relationships between data groups (e.g. the "Facility Site" data group in the Facility Identification Data Standard can be associated with one or more instances of the "Geographic Coordinates" data group.)

Figure 5, below, describes a simple data standard example (the State/EPA agreed-upon Date Standard), which only describes the definition and format for one data element. An example of a complex data standard (the proposed Facility Identification Data Standard), which describes a number of data groups and their relationships to each other, has been provided in Appendix E.

**Figure 5: Simple Data Standard Example of the State/EPA Date Data Standard**

FINAL DATE DATA STANDARD AS POSTED ON THE ENVIRONMENTAL DATA REGISTRY

**DOCUMENT DETAIL**

**Title**: Date Data Standard and Business Rules for Representation of Calendar
**Date EPA Document Number**: Not Available
**Abstract**: This data standard and business rules support the implementation and maintenance of the Agency standard for representation of calendar date. This standard provides for consistent numeric representation of calendar date to facilitate interchange of date data among Agency information systems.
**Purpose**: To layout a data standard and business rules for representation of calendar date.
**Organizational Author**: Alvin M. Pesachowitz
**Version**: 1.0
**Document Date**: 19990120 (YYYYMMDD)
**Access Constraints**:
**Coverage**:
**Coverage Period**:
**Cataloging Source**:
**Create Date**: 19990223 (YYYYMMDD)
**Change Date**: 19990616 (YYYYMMDD)
**Program Component**:
**Expiration Date**: (YYYYMMDD)

**DATA ELEMENT INFORMATION**

**Registry Name**: Date
**Identifier**: 5432
**Version**: 1
**Definition**: A particular day of a calendar year.
**Example**: 19961011

**VALUE DOMAIN INFORMATION**

**Datatype**: Date
**Maximum Character**: 8

## C. Business Case and Critical Features

Implementation of commonly used data standards on the Network where appropriate will help improve data consistency and quality. Wherever possible, DETs will incorporate data standards to bring consistency to the information being shared. Standardization is especially important for information (like facility or location) likely to be integrated with other users' data. If successful, use on the Network of these cross-program standards in DETs may be one of the most significant contributions the Network itself makes in supporting the integration of what have historically been program-specific flows.

## D. Government Issues

The Data Standards Council cannot bind an agency to using a standard. Individual agencies will determine if, when, and how they might use a standard developed under the auspices of the Data Standards Council.

The Data Standards Council will monitor and act as liaison to other parties creating relevant data standards. Some of the standards currently in use were developed by unrelated government agencies. For example, the standard industrial classification (SIC) codes originally developed by the Department of Commerce are widely utilized by many government agencies and are being updated by a group of federal agencies. Various standards are also being developed by industry groups, the American Chemical Society, American Biological Society, the U.S. Fish and Wildlife Service, and interagency groups such as the Federal Geographic Data Committee. Coordinated development through the Data Standards Council will prevent agencies from developing standards that already exist. State environmental agencies that have already developed data standards are encouraged to bring these to the attention of the Data Standards Council and appropriate workgroups to expedite their recognition and use in Network DETs.

Data standards will only prove useful if they are widely accepted and used by the trading partners on the Network. EPA, in approving the use of a DET in fulfillment of a delegation agreement, will likely only approve those DETs compliant with the relevant standards. In establishing DETs for trading partners (e.g., other state or local governments), states may apply similar requirements. While no formal mechanism for enforcing the use of data standards is envisioned, the Network (and participants) should promote and encourage the use of these standards whenever possible.

# 7. Component 2: Data Exchange Templates[4]

## A. Background

Data exchange between environmental regulators to date has been characterized by a series of negotiated agreements to use a specific file format or a specific computer program. The vision for Network exchanges relies on agreed-upon, open, neutral, standards-based data exchange templates for defining and describing the information that is exchanged and on secure Internet transaction protocols for actually moving the information between trading partners. This foundation will allow for adaptability in the shared information, independence for the partners involved in the exchange, and resilience for the specific flow as new technologies emerge.

The IMWG recognized the many benefits associated with information accessibility, including elimination of the requirement for states to load data into national EPA systems (e.g., PCS, AIRS, RCRIS). Use of data exchange templates that are standards-based and technology-neutral will encourage broad Network participation by states, and preserve existing trading partners' internal mechanisms (database software and structure) for storing and managing their information.

Wide agreement is nonetheless necessary on what constitutes acceptable DETs. To understand the definition of DETs in the context of the Network, it is important to distinguish between DETs and transactions (templates containing data.)

## B. Definitions

Figure 6 presents the hierarchy of components relevant to DETs. Each major component is described in the following sections. (Data elements and data groups are defined above, in Component 1: Data Standards.)

### Data Exchange Templates

Data exchange templates identify types of information (data elements and data groups) required or allowable for a particular type of data set according to predefined standards. DETs are empty and contain no data. They simply define the format data must take prior to exchange. DETs will rely on existing data standards where appropriate to increase data quality and consistency. A complete template contains the data groups necessary to describe a specific business event (e.g., issue a permit, initiate an enforcement action.) Figure 7 presents a simplified example of a DET for regulated facility information expressed in extensible markup language (XML).

[4] In May 2001, the WC3 issued the "Schema" as a formal recommendation (its strongest endorsement) for how the structure of XML documents should be expressed. Unless noted otherwise, the DETs discussed below are presumed to be expressed using Schema.

**Figure 6: Data Exchange Template Definitions and Examples**

| Components: | Examples: |
|---|---|
| Data Element | First Name, Facility Identifier, Mailing Address Line 1 |
| Data Group | Mailing Address, Facility Site, Affiliation, Enforcement Action |
| Data Exchange Template | List of enforcement actions taken (conforming to the facility identification and enforcement data standards as developed and adopted.) |
| Transaction | State's enforcement action records in the State/EPA Data Exchange Template format |
| Transmission | State's enforcement action records in the State/EPA Data Exchange Template format submitted to EPA on September 8th, 2000 |

**Figure 7: Blank Data Exchange Template**

```
<StateEpaFacilityExchangeFormat>
    <FacilitySite>
        <FacilityIdentifier> </FacilityIdentifier>      } Data Element
        <FacilityName> </FacilityName>
        <Affiliation>
            <AffiliationType> </AffiliationType>
            <Individual> </Individual>
            <MailingAddress>
                <MailingAddressLine1> </MailingAddressLine1>
                <MailingAddressCity> </MailingAddressCity>
            </MailingAddress>
        </Affiliation> ...   } Additional data elements would constitute the entire data set for FacilitySite
    </FacilitySite>
</StateEpaFacilityExchangeFormat>
```

FacilitySite Data Group

## Transactions

Transactions are defined as a specific set of data exchange templates containing data. Figure 8 represents a simplified example of a transaction containing a sample of state environmental agency regulated facility information. Transactions may consist of multiple instances of a specific data exchange template, each containing data. Information flows over the Network when transactions are exchanged with a trading partner. Transactions will be converted from their electronic format to a human-readable or a different machine-readable format via no- or

low-cost commercially available tools (i.e., a browser). Ancillary documents, such as maps, text documents, reference documents, and images may be carried in their native formats or referenced via URL Web links. Existing XML-based formats are available for all these types of data.

**Figure 8: Example Transaction**

```
<StateEpaFacilityExchangeFormat>
    <FacilitySite>
        <FacilityIdentifier> 011342 </FacilityIdentifier>
        <FacilityName> Zeke's Local Gas Station </FacilityName>
        <Affiliation>
            <AffiliationType> Owner </AffiliationType>
            <Individual> Zeke Brown </Individual>
            <MailingAddress>
                <MailingAddressLine1> 2343 22nd Street </MailingAddressLine1>
                <MailingAddressCity> Tacoma </MailingAddressCity>
            </MailingAddress>
        </Affiliation> ...
    </FacilitySite>
    <FacilitySite>
    <FacilitySite>
    <FacilitySite>
    <FacilitySite>
    <FacilitySite>
    <FacilitySite>...
</StateEpaFacilityExchangeFormat>
```

**Additional FacilitySite Records would be included to complete the transaction file**

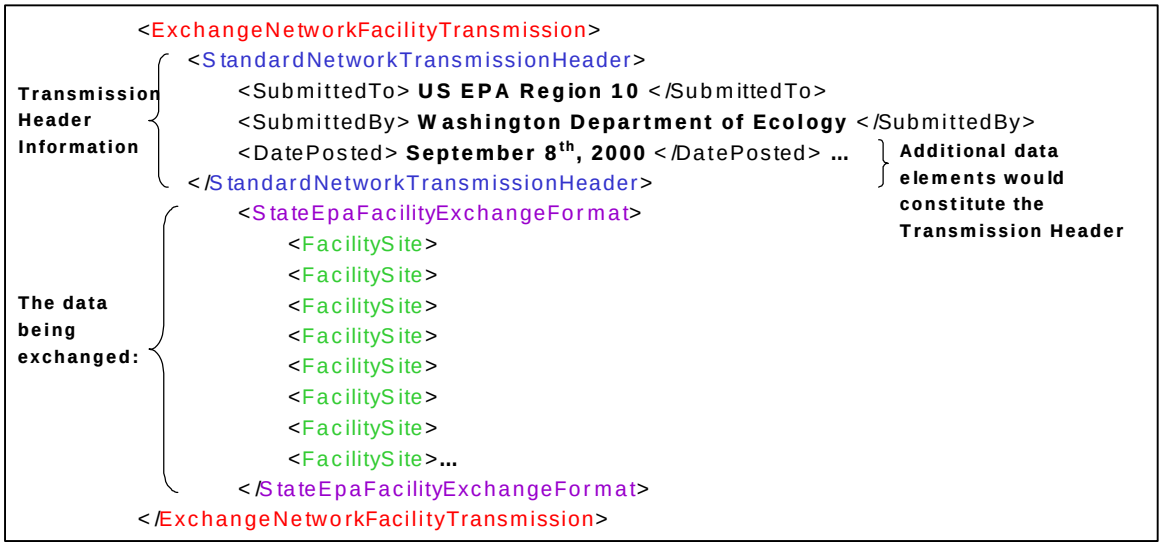
Transmissions
One or more transactions moved across the Network between trading partners constitute a transmission. Figure 9 represents a simplified example of a transmission of regulated facility information from a state agency to EPA.

## C. Business Case and Critical Features

Data exchange templates define data available on the Network. It is assumed that the first series of DETs will support traditional state-to-EPA data flows for the major regulatory activities, such as hazardous waste management, air permitting and water quality monitoring.

The DETs define not only data groups and elements, but also their logical interrelationships. For example, an appropriate DET can make it clear that a facility has one or more permits; each of which has permit conditions. Such a mechanism allows efficient exchange of data without needing agreement on format.

**Figure 9: An Example Transmission**




## Necessary Element of Network Exchange

The existence of an agreed-upon and published DET is a major element in distinguishing Network exchanges from other Internet publication of data. All Network exchanges must conform to a specific DET.

## Maintenance of Data Exchange Templates

Designation and maintenance of a template registry is required. Effective data exchange requires that all trading partners have access to all the DETs being used on the Network. The management of this registry is identified as a core feature of the Network Administration function described in the next section. The choice to use XML as the sole DET language on the Network brings with it a host of tools for the management of these repositories, perhaps most significantly the capability to have the repository referenced in real time each time a DET is used. This means that much of the logistical management of "versions" of DETs becomes automated. When trading partners use a DET, they simply reference the repository and the template is served up. This feature also provides a powerful way to encourage and monitor the use of DETs and the use of standards in DETs.

## Extension of the Business Process

Individual DETs mirror specific business events. At first, one or more specific templates are anticipated for each business process involved in data exchange. Over time, these templates may be merged into a smaller and better-integrated set.

## Format for Data Exchange Templates

The Blueprint Team considered a wide range of options for DET languages, including making no specific recommendation. After much deliberation, including the council of outside technical

experts, the team elected to focus solely on XML as the language for DETs[5]. XML is the best tool for trading partners to unambiguously express and then validate the data they wish to exchange. The team made this decision with the full knowledge that the technology is still immature, and that few existing XML flows exist now.

## D. Government Issues

Data exchange templates can and will be established in a number of ways for use on the Network. For the existing state-to-EPA data flows, DETs will likely be developed by workgroups of state and EPA staff members familiar with those individual flows; a mechanism for joint development, adoption and sharing of DETs may be desirable. As the Network grows (both in number of trading partners and in the amount of information available), DETs will be created as needed by the trading partners. This flexibility will allow the Network to evolve and meet the needs of a much wider set of trading partners.

[5] The Blueprint Team recommends the exclusive use of XML as the common basic interchange language for data flows. The Blueprint uses the term "DET" as neutral term for how that XML would be structured, because the earlier XML structuring language (DTDs) were still in use and the follow-on XML structuring language of Schema had not yet been finalized

# 8. Component 3: Trading Partner Agreements

## A. Background

The electronic commerce most familiar to users of the Internet is the business-to-consumer variety. Typically, a consumer accesses a Web page and is guided by the rules embedded in the application - be it Amazon's shopping cart or some other mechanism. The user interacts live with the application, and may back out if an application imposes unacceptable conditions. For example, if specific personal financial information is needed to complete a transaction, the user may simply decline to submit it, canceling the transaction. Effectively, the website and the user impose conditions and reach agreement through completion or termination of actions.

Increasingly, business and government are seeing the value of electronic transactions that go a step further – electronic transactions initiated by a system owned by one party and negotiated with a system owned by another, without intervention of a user. For example, a business enters a purchase order into an automated system. That system contacts a specified set of vendors, places the order, negotiates details of payment, delivery and terms, and electronically executes the transaction. Prior to implementing such systems, businesses have ensured that such transactions protect the rights of all parties, and that the systems truly reach a common understanding of terms, conditions and other details. At the lowest technical level, considerable agreement is needed simply to begin negotiation of a transaction. In the business-to-business world of electronic commerce, the agreements needed to enable commerce are called trading partner agreements (TPAs). IBM Corporation has developed a standard for such agreements, available in the public domain (Sachs et. al., *Executable Trading-Partner Agreements in Electronic Commerce*). Several of the core elements that IBM includes in its standard and in the electronic language used to express the agreement (TPAmL) for the private sector were adapted to the public sector for this component of the Network.

Similar unattended electronic exchanges of information will be needed for data exchanges over the Network. Much of the methodology emerging from the business-to-business e-commerce is directly applicable to any such transactions. Statutory oversight requirements, negotiated agreements between states and EPA and mandatory reporting requirements introduce additional conditions unique to government or specific to the environmental information Network, which are discussed below.

## B. Definition

Trading partner agreements are documents formally adopted by two or more partners for the purpose of defining the responsibilities of each party, the legal standing (if any) of the proposed exchange, and the technical details necessary to initiate and conduct electronic information exchange. TPAs may apply to exchanges initiated by the sender ("push" systems) or those initiated at the request of the receiver ("pull" systems). TPAs are necessary when automated exchanges are to take place without operator intervention if the exchange is intended to meet or replace any mandatory reporting requirement. They are advisable between any parties (e.g., states) who wish to establish an ongoing business process involving automated electronic

exchange of information. Specific agreements regarding electronic data exchange between EPA and the states, as currently included in PPA and SEA documents, exist as the current implementation of TPAs. Future TPAs may take the same form, be drafted to complement a PPA and SEA, or stand alone. TPAs do not apply to one party's access of data provided through a public access site. Such access may be negotiated when both parties agree they wish to exchange data.

## C. Business Case and Critical Features

In practice, a TPA may be lengthy and highly technical, or relatively simple, based on the needs of the specific data flow and the existence of other governing documents. In the more simple case, a flow or exchange may have already been defined as part of a PPA, SEA, Memorandum of Understanding (MOU), or other agreement specifying timetables, data requirements, terms of governance, technical specification and details of flow mechanisms. The TPA for a data flow may then contain only the basic elements missing from the original agreement and reference to these other documents. Such robust technical or governmental frameworks may not exist in other types of flows and state systems. The TPA for a data flow would then need to be more comprehensive and detailed.

The following items should be addressed in the TPA, either directly or by reference to another document:

- **Parties**. This section identifies the organizations involved in the TPA and describes the general purpose of the agreement.

- **Legal Framework.** This section includes governance, standing and applicability issues that apply to the partners. The TPA should address the effect of the agreement on other interparty obligations. For example, it needs to address any reporting requirements met by the agreement. The TPA should also address applicability to all levels of participating organizations.[6]

- **Security.** This section identifies the level of Network security to be used and the specific parameters such as certificates used for authentication, non-repudiation and digital envelope, and other security issues.

- **Data Definition.** This section describes the specific format and structure to be used for exchange and the URL of record for the format.

- **Communication.** This section specifies the transport protocols and electronic addresses of the parties.

- **Message Exchanges.** This section discusses rules for submitting and responding to requests for data and the timing of data exchange. It includes a list describing the requests that parties can issue to each other. These actions are the independent units

[6] If executed by a Region and a state, the relationship to EPA Headquarters requirements must be addressed.

of work. The action definitions reflect the associated message flows between the invoker and the service provider, responsiveness, failure handling and other attributes. This section should address the expected update cycle for data of record (e.g., the steward agency will enter data within five business days).

- **Definition of Roles and Responsibilities.** This section outlines specific roles and requirements of parties related to performance, reliability and use of data.

  - Internal Systems Requirements. The TPA does not address partners' internal computer systems unless the electronic exchange is predicated on maintenance of specific internal requirements (e.g., EPA's proposed electronic reporting rule). In such cases, they should be specified.

  - Performance and Reliability. The expected availability of participating systems is specified here.[7] For high-volume systems, the TPA should also identify system performance expectations (e.g., transfer speed, response times).

  - Exchange Failure. Because some exchanges may be mandatory (once voluntarily included in the TPA), the TPA should identify actions required by each party should the exchange fail.[8]

  - System Failure. When the exchange is intended to duplicate data locally, the TPA should address initial synchronization of participating databases and recovery following system failure.

  - Quality and Stewardship. The TPA should specify the definitive source for shared data. The TPA should outline expectations regarding timeliness of data entry, error detection and correction, and other conditions upon which acceptability of the data is predicated. [9]

  - Use of Data. Intended routine uses of the data are specifically addressed to the extent needed in order to understand the responsibilities of the parties. Generally, the allowable uses of data need not be included in a TPA, as the data would be reported by some means in any case. Once delivered, the receiving party is still bound by such considerations as confidential business information (CBI) or enforcement-sensitive data, as if the data had been exchanged in the traditional manner. The TPA may need to address how such data, if mixed with other data, will be identified. If one party wishes to exclude a specific use that would

[7] Situations are already arising where external data is included in public access products. Linking of significant portions of another's web products may be reason to execute a TPA indicating that there is some agreement to maintain specific content at a specific location.

[8] System error response procedures are a part of communications protocols. This item is intended to address business continuation in the event of failure.

[9] "On demand" data exchange introduces these factors. Periodic reports are expected to be complete for the period covered. Where these are replaced by ad-hoc sharing of data, the trading partners need an understanding about the condition of the data on an ongoing basis.

otherwise be enabled by the exchange, it should be addressed. For example, in providing non-mandatory data, states have indicated in a PPA that EPA may not use the data for program evaluation.

- Dispute Resolution. The agreement describes procedures for settling disputes related to the terms of the agreement.

- **Parallel Paper Transactions.** Any expectations for exchange of documents on paper in addition to electronic format for a portion of or the entire duration of the TPA are outlined in this section.

- **Record Retention.** This section addresses issues surrounding transmission logs and requests for historical data.

- **Duration.** This section identifies the period of time for which the agreement will remain in effect.

- **Termination.** This section specifies conditions for termination of the TPA as a whole, including written notice and the effect of termination on other rights and obligations.

- **Addenda.** This section describes if and how addenda may be added to the agreement.

The TPAmL Schema noted above, as well as other TPA templates, provide a structure and format for expressing many of these conditions as the Network begins. Other initiatives, such as ebXML (e-business XML), are basing their efforts on the TPAmL work. It is likely that initial Network flows will employ a variety of TPA formats. As best practices emerge, they can be codified by the Network administrator into TPA templates. In addition, EPA or other major trading partners may establish templates as a starting point for TPA development.

## D. Government Issues

A very important feature of many data exchange (especially e-commerce) networks is that they are bilateral (or peer-to-peer) and therefore self-enforcing. For example, the Internet itself, at any point in time, is simply the collection of computers that have agreed to route TCP/IP among each other. When users sign onto their Internet service provider (ISP), or when that ISP links to its ISP upstream, they agree to use TCP/IP and abide by a basic user agreement. If users are not willing to use the TCP/IP standard, they cannot connect. If they violate their user agreements, their ISP will turn them off. E-commerce networks operate in a similar way. Sony and IBM execute a TPA and begin exchanging messages. If Sony sends the wrong part, or misrepresents a catalog entry, IBM deals with Sony; the e-commerce "administrator" (i.e., RosettaNet) does not become involved. Thus the larger network polices itself without the involvement of a central authority.

It is envisioned that the Network will be governed by bilateral TPAs and supported by a basic "Network User Agreement" agreed to by all partners when they join the Network. The Network User Agreement will define basic terms and conditions for participation in the Network. The agreement will be common to all Network data flows and will not need to be negotiated separately for each set of trading partners. Any special terms not included in the overall agreement will be included in separately negotiated trading partner agreements when determined necessary.

When a party attempts to provide data that either does not comply with the agreed-upon exchange format or does not meet some other term of the TPA, its partner is in a position to respond using its available authority. That data should not become part of the Network. If the data meets the requirements of agreement, it becomes part of the Network in good standing. By making the TPA explicit about data quality, the Network attempts to establish some baseline for the reliability and trustworthiness of its data. The use of XML provides data originators with significant ability to "self-validate" their own transmissions and recipients with the capability to assess the conformance to the DET.

Unlike engaging in commerce or running the Internet, the purpose of the Network is to support the flow of high quality environmental data. Not all of this data is, or will be, covered under a bilateral TPA. Once the Network is established, members may wish to make a form of quality declaration for given data on their node. For example a state may wish unilaterally to declare a given data source as its "official source of record for the state field burning program." Such a declaration would explicitly document the pledge of the participant to establish and maintain a specific data source as if there were a vigilant trading partner. This declaration would be similar in format and content to the standard bilateral template. It might even include a "complaints" section where data users could contest or otherwise comment on the data. Eventually, the Network administrator or others could fulfill some kind of "audit" function for these data sources, perhaps codified in a TPA, as service to members who choose to offer this type of information. (This external audit function is a feature of some e-commerce networks.) This function would be analogous to a Certified Public Accountant (CPA) auditing a firm's financial statement as accurate. As in the case of the bilateral TPA, the objective of this formal statement would be to provide a baseline of reliability and credibility to Network data.

# 9. Component 4: Technical Infrastructure and Network Administration

## A. Background

The technical infrastructure of the data exchange Network will use the Internet in the same way as many private e-commerce initiatives. Open standards (i.e., standards that are not tied to a specific technology or vendor) will be utilized whenever possible to encourage information sharing. The proposed infrastructure is a "front door-to-front door" framework. The only technology decisions that are being discussed operate on the actual exchange of information between partners and do not deal with the internal workings of how an agency manages and stores its information. These decisions will focus on transfer mechanisms and data exchange formats, which are the two key technical areas that relate to actually exchanging information between trading partners. Because of this, there will be no significant impact on the technologies that an agency chooses to use for database design or application development. The investments and decisions that agencies have made and continue to make concerning internal storage and management of information will not be affected. Also, because the technology infrastructure of the Network will be based on open standards, participating agencies will have tremendous flexibility in choosing hardware, software and service providers to implement the Network-specific technologies that will be needed to fully participate.

## B. Definition

The technical infrastructure of the Network is the software, hardware and protocols used to make it function. This Blueprint identifies the following elements of this infrastructure:

### Basic Network Protocols

All information exchange on the Network will occur utilizing the following protocols:

- Transmission Control Protocol/Internet Protocol (TCP/IP) – communications protocol used to connect hosts on the Internet. TCP/IP is the de facto standard for transmitting data over networks.
- HyperText Transfer Protocol (HTTP) – protocol used to define how messages are formatted and transmitted and what actions servers and browsers should take in response to commands.

### Languages for Expression and Construction of Data Exchange Formats

All of the data and all DETs on the Network will be expressed in Extensible Markup Language. (XML). XML is a language for the creation of Web documents and forms. It facilitates the definition, validation and interpretation of data between applications and organizations.

### Request, Transmission and Query Protocols

Initial Network flows may use only the simplest possible request/acknowledgements for transport between nodes. In some cases this may be a simple "get" or "post" command in HTTP. The ability of a node to respond to predefined queries, constructed on the basis of DETs, is a powerful but more advanced capacity that will develop over time. Many competing protocols are in development for these kinds of functions; they include SOAP (Simple Object Access Protocol) and XQL (Extensible Query Language). First-generation Network exchanges may be able to use much simpler subsets of these tools as a common starting point. In addition, several broader proposals, such as ebXML (www.ebxml.org) may address both the DET and request/transmission protocols as well as other Network components. As experience is gained in implementing these approaches, and as the approaches themselves mature, they can be standardized and coordinated by the Network administrator.

Limiting queries to those prescribed with the DET will allow node managers to ensure that they can be easily serviced.

### Security (see table and section below) (sHTTP, SSL and PKI)

- Security – techniques for ensuring that data being transmitted or stored in a computer cannot be read, altered or compromised by those not intended to do so. This will include technology such as Public Key Infrastructure (PKI) that verifies and authenticates the validity of any information Network partners involved in an information exchange.

- Secure Socket Layer (SSL) – the connection over which a protocol that uses a private key to encrypt data is transferred. SSL is supported by both Netscape Navigator and Internet Explorer and can be used to transmit any amount of data securely. URLs for Web pages that require a SSL connection start with a "https".

- Secure HTTP (S-HTTP) – protocol for transmitting individual message securely.

## C. Business Case and Critical Features

The technical infrastructure of the Network will be based on the small set of core technologies identified above. As in the example of the WWW itself, some technologies will be required while others will present an evolving menu of specific options. It is anticipated that the Network will define several levels of security (described in Table 1), available to trading partners as needed. The specific level of security to be used for a given flow would be documented in the TPA, although the tools to implement the agreed-upon security level would not.

| Table 1: Description of the Four Network Security Levels | | |
|---|---|---|
| **Security Level** | **Characteristics** | **Approach** |
| **Level 1** | Public information that requires no authentication or certification of integrity. Like all Network information, this information is protected from unauthorized modification at its node. | This information will be available through the Internet on a public, non-secure website. Information can be transmitted without encryption or special security measures. |
| **Level 2** | Information that requires some additional level of authentication (i.e., that it is the State who is submitting the data) and a higher level of integrity protection. This data may require some level of confidentiality. | This information will be available through the Internet on website that is secured using Secure Socket Layer (SSL). The use of SSL allows the users to authenticate that the site being accessed is an approved environmental agency website, and provides privacy by encrypting all data in transit. SSL also provides data integrity protection. |
| **Level 3** | Information at this level requires bi-directional authentication and a higher level of confidentiality. All data submitted by users to environmental agencies is to be treated at this level or higher. This data is of a highly sensitive nature passed between agencies but does not require digital signature. This level can apply to person-to-person and server-to-server transactions. | Access to this information is protected by SSL at the server level, and by the requirement for users' digital identity credentials. These credentials will be in the form of X.509 version 3 digital certificates issued by a Public Key Infrastructure (PKI) that the environmental agency determines meets a sufficient level of assurance in identity proofing and credential protection. Once users have been authenticated, they will be permitted to access only that data to which they are allowed. |
| **Level 4** | Information protection that requires non-repudiation in addition to privacy, authentication and data integrity. Generally, this information is the electronic version of current paper processes that require an ink signature. This information may be in the form of data going from the agency to external users, or may be reports, applications or other information going from external users to the environmental agency. | This information will be protected by requiring a digital signature "affixed" to the data that can be validated at the time of acceptance of the information by the environmental agency or the external user. Digital certificates issued by an approved PKI will be used for digital signature. |

| Table 2: Technological Characteristics of the Four Network Security Levels | | | | |
|---|---|---|---|---|
| **Security Level** | **Standard Internet Firewall** | **Secure Socket Layer (SSL)/Authenticate Originator (Digital Certificate)** | **Authenticate both Trading Partners (Digital Certificate)** | **Digital Signature Affixed** |
| **Level 1** | Yes | -- | -- | -- |
| **Level 2** | Yes | Yes | -- | -- |
| **Level 3** | Yes | Yes | Yes | -- |
| **Level 4** | Yes | Yes | Yes | Yes |

| Table 3: Public Sector and Private Sector Examples of the Four Network Security Levels | | |
|---|---|---|
| **Security Level** | **Commercial Examples** | **Environmental Agency Examples** |
| **Level 1** | CNN.COM | Public viewing of ambient environmental conditions |
| **Level 2** | AMAZON.COM Ordering Process | List of certified state laboratories |
| **Level 3** | Transmission of supply and order information between trading partners. | State submission of a formal report required by EPA |
| **Level 4** | Contractual binding documents and e-mails | State submission of formal report to EPA which requires an official signature |

These levels were developed on the basis of technologies states and EPA are already implementing. EPA and states provide significant information on their websites under Level 1 security. Many states have already established Level 2 security for their commerce functions. Levels 3 and 4 represent combinations of these and will be piloted as part of the Central Data Exchange Action Team chartered by the IMWG. Because they are based on open standards, it is likely that members will use a variety of technical architectures to establish the security of their nodes behind various levels of firewall. XML data travels over the same portion of this infrastructure as web pages, and with the explosion of XML use, security measures anticipating these architectures are readily available. In many cases, these features are built into the server/e-commerce software currently in place.

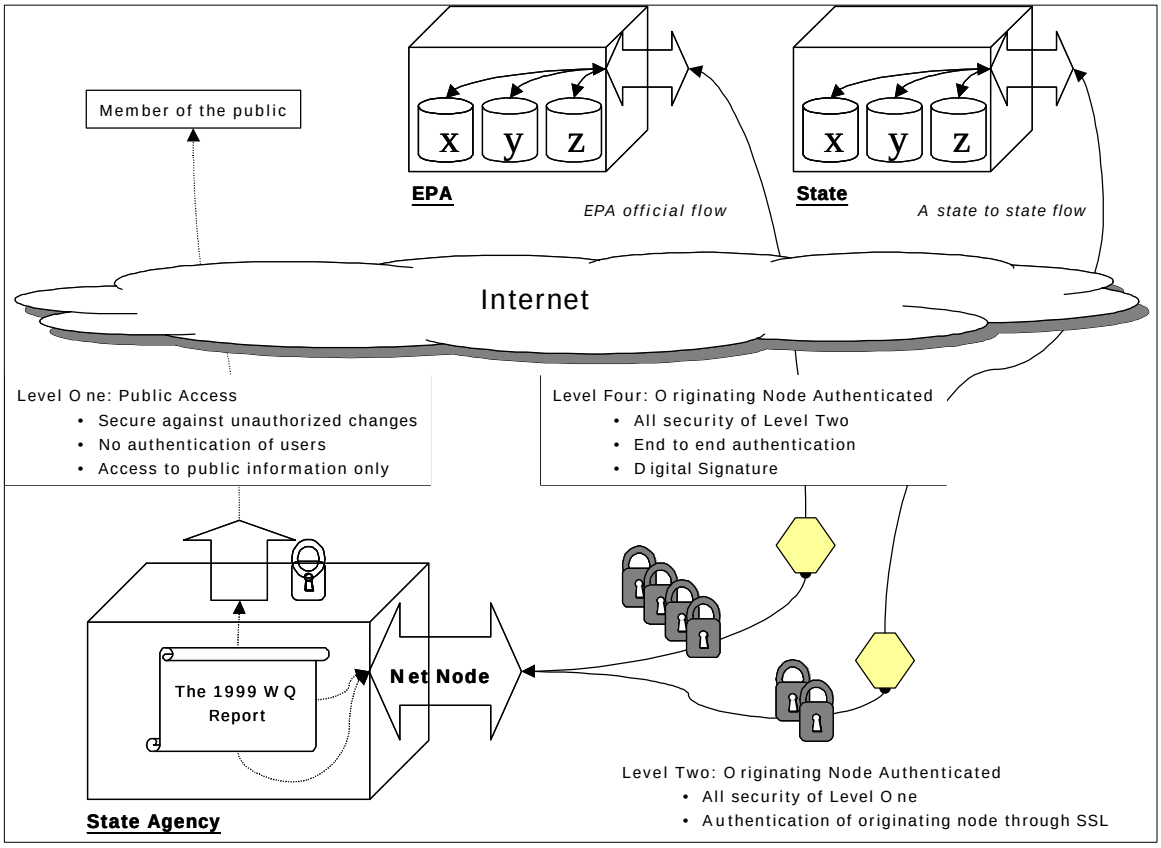
Most of the information on the Network is anticipated to be public. Certain transmissions of this data (i.e., those constituting official intergovernmental flows) will require a given security level, but the same data may also be available via the data originator's Network node (and perhaps their public access website) at a different security level (e.g., Level 1). As depicted in Figure 10, the identical information may flow from the node under different security levels depending on the partner. The ability to manage these relationships will be a significant portion of the administrative and technical costs of running a Network node. E-commerce software (e.g., Microsoft's Biz-talk server, WebMethods, or Mercator) fulfills this function.

This approach is based on the following additional observations/findings:

- Internet security is an issue agencies will increasingly confront whether this or any other Network evolves.
- Agencies will have to face enterprise (e.g., Network node) security issues as they move to the conduct of business and protection of their websites.
- All agencies have to manage the traditional more intrusive relationships they have with trading partners. Many agencies are attempting to minimize these types of interactions to reduce the burden on staff and resources.
- Many of the Network security features discussed here represent significant investments, but they are investments that will be required by any agency wishing to realize the benefits of moving into the Internet age and participating in any form of e-commerce. They offer a great opportunity for synergy and cost savings by allowing Network members to focus on securing a single enterprise Network port rather than an ad-hoc collection of individual feeds and services. Implementation of the Network

could reduce the burden placed on state and EPA information technology (IT) personnel by reducing the number of systems required to communicate with the various EPA programs.

**Figure 10: A given information request may flow over the Network under various security levels**




It is even possible that the Network approach could reduce the security risks associated with some data flows by establishing standard protocols and technologies. The more diverse and non-standard the data flows, the greater the security exposure. For example, EPA faces a challenge of providing direct client access by state staff to its national systems. This usually involves EPA developing a piece of software that a state agency uses for the specific purpose of access to and uploading of information to EPA. These states are "clients" to EPA servers. This approach also requires a separate, secure transport mechanism between EPA and each state using the software. This is usually accomplished by setting up a file transfer protocol site for each trading partner, which increases the need to manage multiple security relationships. While some states will require this level of access for the foreseeable future, many of the flows that currently require this type of access might be migrated to the Network. The need for direct access to EPA systems could be reduced in such flows, in that EPA would initiate data requests from secure state servers, states would access information on EPA servers via the EPA node, or both. This offers the potential for dramatic simplification of EPA's security predicament by limiting the number of external clients with direct access to its servers. This scenario is also consistent with EPA's decision to focus its enterprise e-commerce flows through the CDX facility.

## D. Government Issues

It is critical that the Network remain vendor neutral and flexible. The goal of the Network is to encourage information sharing and to reduce the burden on participating organizations. Use of a particular software or hardware technology cannot be required to participate in the Network.

## E. Introduction to Network Administration[10]

The technical infrastructure section above outlines the specific tools and technical standards (e.g., XML, HTTP, SSL) proposed for the Network. Like a local area network (LAN) in agency offices or the Internet itself, the Network will also require a minimal (but critical) administrative capability. A Network administrator would not take the place of lower level technical standards bodies (like the Internet Engineering Task Force (IETF)), the TPA or high level intergovernmental agreements. Like a LAN administrator, a Network administrator would establish recommendations for how to use the Network, not what data to access or what to do with that data. For example, it is expected that states and EPA will use the Network to replace system-dependent flows of data under delegated programs. EPA will use this data as part of its oversight of national programs. The Network administrator will simply support the flow of this data, not its use in oversight. From the administrator, one could learn how to get the status of a facility's permit from a state node, but not whether that status is appropriate or timely.

### Parallels In the Public and Private Sector for Network Administration

A Network administrator would undertake whatever functions are needed to support the Network that are not best done by the individual participants acting alone or with their individual trading partners.

- Provision of basic reference information about the Network, its participants and their data.
- Maintenance of a repository for DETs, transaction protocols and trading partner agreement templates, registered on a voluntary basis for participants' reference and use. This registration may be a requirement for a given TPA.
- Maintenance of a repository for TPAs registered on a voluntary basis that result in new data sources at a member node. This registration may be a requirement for a given TPA.
- Provision of minimal "steering group" guidance.

Given the difficulty, expense, and slow pace of wide-scale collaborative change, only the absolute minimum required to initiate the Network is proposed. Many participants will identify scores of other functions for a Network administrator (e.g., maintain a well-indexed search engine, build a value-added portal that links the participants' sites, rate the quality of the data on member websites, provide technical assistance to members). Such ideas can be considered by the IMWG after the basic infrastructure of the Network is established. Furthermore, many of these

[10] See Section 14: Network Administration Report to IMWG.

activities can be done by EPA working with states, by groups of states working together, or by the IMWG itself.

Specifically, the following broader functions that might be performed by a Network administration were considered but deferred:

- Identification, prioritization and sponsorship of DET creation.
- Active promotion and expansion of the Network membership.
- Development of readiness assessment guidelines for potential trading partners.
- Development and distribution of a "quick start" kit that allows partners to participate.
- Shared use and support of an expert team to conduct readiness assessments, and setup of a partner site with "quick start" kit.
- Maintenance of a list of lessons learned and frequently asked questions (FAQs) (including node security).
- Establishment of a "test bed" facility to be accessed and used by all partners while developing new transmissions.
- Development of Network performance metrics.

The IMWG may elect to begin work immediately on these functions, but such efforts should be independent of those supporting the minimal Network administrative capability identified above. Some of these functions may be appropriate for immediate EPA sponsorship. They are discussed further in the Member Organizational Infrastructure Section.

## By Whom Would This Function Be Fulfilled?

No single entity governs the Internet or the WWW. The Internet *is* agreements to use common technologies and standards; that is all. The closest things to governance are groups that perform very specific and limited registration functions (often private sector firms that compete with each other) and groups like the W3C, a non-profit consortium that develops "standards" for infrastructure like HTML or XML. None of these groups has legal authority to force people to follow rules.

When people use the Web to conduct business, the technology and the governance of the Internet itself is transparent. This is the target for the Network as well – that participants simply grab, adapt and use the tools offered by the Network to create flows between trading partners. Like an Internet e-mail, individuals do not depend up on their ISP or the W3C to tell them what they can or should put in their business correspondence, or how to handle any aspect of a debate that might arise as a result of the message itself.

## How Would This Function Be Fulfilled?

Models for various aspects of Network governance and administration are listed in Table 4.

| Table 4: Models of Network Governance | | |
|---|---|---|
| Area | Governance Bodies | Functions |
| Internet Infrastructure | IETF (Internet Engineering Task Force)<br><br>Registrars | ❑ Create technical recommendations for underlying technology (like the format of email messages, and internet addresses like 207.18.19.166 and their domains "www.ibm.com"<br>❑ Private firms that are authorized to register domain names and addresses. |
| WWW Infrastructure | W3C | ❑ A non-profit consortium that develops underlying technologies of the web like HTML and XML. Issues formal recommendations. |
| Visa | "Visa" association | ❑ A membership association of banks and merchants who agree to abide by "Visa standards" for transactions. A steering group sets technical and performance policies (e.g., you must accept/process Visa charges using a standard transaction set or you don't get to use the Visa network). There is no one Visa corporation - just a holding company whose shares are all owned by members. |
| RosettaNet | RosettaNet–RosettaNet is a non-profit consortium of companies dedicated to e-commerce tools and standards. | ❑ Develop RosettaNet technical and process standards for use by members. |
| OASIS | OASIS- (Organization for the Advancement of Structure Information Standards) | ❑ A non-profit organization that supports members in development of standards<br>❑ Is host to both TPAmL and ebXML in partnership with the UN |

After review of these and other models of administration, and much debate, the Blueprint Team proposes the following summary principles for creation of the Network administrative function:

- It will be kept to the absolute minimum needed to start the Network, and expanded to provide more functionality as it becomes credible to do so.
- It will focus on the core tasks of voluntary registration of members, member node catalogs, TPAs, DETs and query/request protocols. It will host a simple website with reference information about the Network.
- It will have some independence from individual members. This means it will not be solely administered by any state, by EPA or by ECOS. It may have some third party standing.
- It will remain independent from the efforts of participants to promote and expand the Network.

This version of the Blueprint does not offer a final recommendation on the specific administrative structure that should be implemented. Instead, the Blueprint Team requests the IMWG authorize an extension of its charter for a short additional period, after the IMWG meeting to prepare a specific recommendation. **However, this is by no means a reason to delay any aspect of establishing Network flows. We stand to learn much by doing so immediately.**

## Government Issues

Aside from the issue of how the Network administrator function is structured, there are few government-specific issues with this component. In most cases these technologies can be used as is because their function is mostly mechanical. By definition, Network administration will NOT include inherently governmental functions.

# 10. Component 5: Member Organizational Infrastructure

## A. Definition

Member organizational infrastructure defines the roles and responsibilities required for Network participants. The term infrastructure is used because these roles and responsibilities will require investment to build, and when effective, should be relatively transparent. Because states and EPA will take the first steps towards implementation together, this section focuses specifically on their near-term roles and responsibilities. As the Network is expanded to other participants (such as tribal governments), their roles and responsibilities will need to be defined as well.

## B. Background

### Purpose of This Section

The preceding components of the Network Blueprint provide the "plumbing" and "electrical specifications" for moving data and administering the Network. Aside from the discussion "What is a Node Really?" above, the components have addressed what the Network looks like from the front door (or node door) *out*.

This Network component focuses on the infrastructure needed to get EPA, states, and eventually other partners "interested, authorized and able" to participate in the Network for their business. It suggests members' internal roles and responsibilities for operating their nodes and supporting (not administering) the Network itself. This section is among the most important in this document; it is also the most preliminary. Because many of these concepts apply both to EPA and states and because they are all interrelated, there is significant redundancy in the current draft. After the workgroup has debated and clarified some of these issues, and as Network flows begin, the details of the roles and responsibilities outlined here will be further refined and documented. Specifically, this section is offered to frame the IMWG's consideration of the Network and its role in supporting the Network. However, none of the issues debated here should preclude two parties from immediately using other concepts in this document to create Network flows between them.

In addition to the basic organizational infrastructure needed, this section discusses what can be done to increase the ***capacity*** of states and EPA to fulfill these responsibilities. It provides a framework for what states can do for themselves, other states and for EPA; it also describes the complex but critical opportunities EPA can take to increase the capacity of states to build and participate in this Network. As this document makes clear, this Network is fundamentally decentralized; yet EPA plays a critical role. While many Blueprint Team members believe this Network (or something like it) will arise with or without EPA's participation, all believe that the important things will happen better and faster if EPA is in at the ground floor. An extensive set of specific options and actions were originally developed as part of EPA's Information Integration Initiative (I-3). Key milestones from EPA's I-3 have been included here. These milestones clearly reflect EPA's public and specific commitment to the Network.

### State and EPA Roles and Responsibilities

State and EPA roles and responsibilities for data exchanges are embedded in a complex, historical web of formal and informal agreements. These include program delegations, annual SEAs, PPAs, PPG and program- or Region-specific agreements. These agreements often overlap, involve different levels of each organization, and in some cases conflict. Worse, in many cases roles and responsibilities are ambiguous, with no one accountable for end-to-end data quality. In other cases, stable program-specific arrangements have developed that include agreed-upon metrics for performance and data quality. This wide variety of experiences and problems makes it easy for participants from different programs and states to hold different opinions on the effectiveness of the existing data exchange system. For parties with stable, negotiated formats and expectations, the Network offers an economy of scale and a refined set of technical tools; for partners mired in ambiguous, conflicting agreements, it presents the challenge of making their obligations and metrics explicit, but also offers the tools (especially the TPA) to do so.

As described above in the TPA component, most of these flows are currently described in terms of obligations of states to feed one or more program-specific EPA information system. The Network will simplify and clarify data exchange roles and responsibilities through the use of TPAs. Each TPA will identify the trading partners and respective node addresses; define the purpose and content of the data exchange; and define expectations for data and transaction quality, security, integrity and frequency. Network participants will need to consider their own requirements for populating internal business applications when developing data exchange templates and TPAs. However, TPAs will not be used to specify how this integration is to be accomplished. The Blueprint Team expressed a strong desire to focus TPAs on business events and processes and on the necessary supporting data and not to constrain the design of DETs and TPAs with the idiosyncrasies of existing internal business applications.

## C. Discussion

By agreeing to host and exchange data on the Network, each trading partner, as a Network partner, assumes and accepts certain roles and responsibilities. These roles and responsibilities will include the following:

### Role – Node Administrator

The Node Administrator, similar to a Web or systems administrator, will be responsible for:

- Software development and implementation (e.g. security, XML)
- System documentation
- Hardware and software maintenance
- Policies and procedures (e.g., security. documentation, change management, problem management)
- Backup and recovery

### Role – Data Steward

The Data Steward, similar to a data administrator, will be responsible for:

- Documenting data and data relationships
- Developing data definitions and data naming standards
- Developing standard calculations and derivations
- Defining data security and retention requirements
- Developing DETs
- Mapping data sources (e.g., business applications) to DETs
- Monitoring data quality

### Role – TPA Administrator

The TPA Administrator, similar to a contract administrator, will be responsible for:

- Developing and approving TPAs
- Monitoring compliance with TPAs

As stated in the Introduction above, effective stewardship of the Network is considered fundamental to the idea of the Network and to its success. The roles and responsibilities described above are considered essential for effective shared stewardship of the Network.

## D. Business Case and Critical Features

The following section outlines some key roles of states, EPA and the IMWG in five distinct areas:

- Supporting the Network Administrator and other shared infrastructure
- Establishing EPA's capacity to build and manage its node
- Establishing EPA's capacity to establish and manage flows with states
- Supporting individual states' capacity to build and manage its node
- Supporting individual states' capacity to establish and manage flows with EPA

These areas are considered from the perspective of EPA, states and the IMWG.

### EPA and State Support of the Network Administrator and Other Shared Infrastructure

As indicated in Component 5: Technical Infrastructure and Network Administration, this Blueprint does not propose a specific structure and seat for the Network administrator. These arrangements will be developed once the IMWG has endorsed the concept of the Network and considered the Blueprint recommendations. Nonetheless, the following roles and responsibilities in supporting this function are clear:

- EPA and states will need to support the IMWG in identifying and establishing the Network administrator.

- State support of this function will likely consist mostly of cooperation and encouragement. States may also be able to contribute direct technical and management resources (staff or expertise) needed to launch this function.
- EPA will also have a special role in presenting and supporting its priority data flows for DET creation, to the extent that this involves the Network administrator.

As the division of responsibilities becomes clear between the Network administrator, IMWG, EPA and states, several additional capacity-building steps could be taken:

- Development of a readiness assessment guideline for potential trading partners.
- Development of a "quick start" kit that allows partners to participate.
- Shared use and support of an expert team to conduct readiness assessments, and set up a partner node with "quick start" kit.
- Maintenance of "lessons learned", FAQs, etc. (including node security).
- Establishment of a "test bed" facility to be accessed and used by all partners while developing new transmissions.

Other support activities were mentioned during development of the Blueprint but have been omitted here for clarity and because their consideration may be premature before the IMWG has discussed the broader Blueprint design.

## EPA Organizational Infrastructure

Early on, the most important opportunity and challenge for the Network will probably be the credible participation by individual EPA staff at various levels in creating flows with their state counterparts. This task will be more difficult for EPA than for states because of EPA's broader, more diverse and more complex data needs and its multiple state clients. Support from states and the IMWG will be needed. EPA's CDX staffs have a direct link to these in-reach efforts through the IMWG Action Team (CDX Action Team), but most EPA data exchanges remain system- and program-specific. Program offices working on current CDX pilots have already begun the Network-oriented data exchange process; but what of the Regional staff person who first hears of these ideas from an eager and aggressive state Chief Information Officer (CIO) who wishes to negotiate something called a TPA? How will *that* person be supported, or at least not stymied? As EPA begins to develop policy and infrastructure to support the Network, it must also ensure that smaller projects succeed. Early Network flow projects (those sponsored by the IMWG and those that arise spontaneously from individual state-EPA initiatives) will form the foundation for later growth.

Under the Network, states and other partners would make their information accessible to EPA's Central Data Exchange facility. EPA would manage its copies of this data (i.e., in the near term, loading data into the existing national systems). While reengineering its systems in concert with EPA's ongoing integration effort, each program will need to help develop exchange formats for its business subject matter area, coordinate with CDX to receive newly retooled transactions based upon these formats, and have the capacity to exchange data in its own system with CDX. These are significant but tractable technical tasks; the real challenge is to manage the following types of change in internal roles and responsibilities:

- Existing programs, policies, processes - Existing delegation agreements that specify information requirements, certain National Environmental Performance Partnership System (NEPPS) agreements, electronic reporting trading partner agreements and informal ad-hoc data acquisition arrangements will all need to converge into documented Network trading partner agreements. These agreements will require the coordination of many people.

- Commitments to conduct business through the Network - Having committed to conducting business through the Network, EPA will need to ensure that its individual programs and regions are able to do so (e.g., have adequate funding and other resources).

- Coordination with internal integration - In addition to retooling information exchanges, and thus system capabilities, EPA is also establishing an enterprise architecture basis for its internal integration efforts. A coordinated, balanced approach may constrict EPA's capacity to retool existing incoming information flows towards the Network vision.

- Central Data Exchange - An operational node on EPA's CDX is required to receive Network data and handle different transmission and exchange formats (transaction sets). Priorities, implementation and resources for CDX development must be established and aligned. CDX and program systems must have the capacity to exchange data. Programs must understand their roles in Network participation, and have the expertise to redevelop their existing information exchanges.

- IT/IRM Policy –EPA's standing IT/IRM policies must be reexamined to determine what is needed to support the Network, concurrent with the reassessment of policies for internal integration and architectural realignment purposes.

- New programs, policies, procedure - A proactive means of handling new laws affecting the Network (e.g., Cross-Media Electronic Reporting and Records Rule (CROMERR)) must be developed. As with other information management concerns, getting in front of the regulatory development process will help create reform that can adapt to future changes.

- Role and responsibilities of Regions - Much of the burden of establishing the Network will fall on the EPA Regional offices. Processes and procedures will have to be harmonized to ensure national consistency. Regions play a central role in the management and organization of their states' TPAs and relationships to NEPPS and other negotiated agreements. Regions can participate with their states to build capacity and extend the Network.

Much of the preliminary planning for I-3 was conducted in parallel with the Network Blueprint work; however, EPA's investment plan for I-3 was due prior to the completion of this document. Because of their direct relevance to the Blueprint, the Exchange Network Infrastructure and Partner Assistance milestones have been included here for discussion purposes only.

**Table 5: EPA Milestones for the Network**

| EPA Internal Integration | Exchange Network Infrastructure | Partner Assistance |
|---|---|---|
| **2000**<br>1. EPA establishes the Information Integration Initiative in support of this vision.<br>2. EPA makes commitment to internal information integration and begins to realign internal structures and resources in support of I-3.<br>3. The utility and expanded opportunities in the use of integrated information to environmental protection programs is clarified via FY2000 demonstration projects. | **2000**<br>1. EPA and state environmental agencies commit to developing a national environmental information exchange network with other partners in environmental protection. A vision of the Exchange Network is documented and supported by ECOS and participating states.<br>2. The State/EPA Information Management Workgroup takes the active lead in developing this vision and Exchange Network. | **2000**<br>1. Utilization of the One Stop Network of state officials in defining the vision and determining partner needs. |
| **2001**<br>1. Initial scope of the I-3 project is refined, well-defined and operational. EPA has made a stated commitment to coordinating its internal integration efforts with the evolution of the Exchange Network partnership.<br>2. A target Enterprise Architecture is in place for EPA's mission functions and is the guiding principle for IT investment decisions and framework for systems development and modernizations efforts.<br>3. EPA Programs and Regions have launched Information Strategic Planning (ISP) exercises and have realigned systems development plans to include utilization of corporate data services and functions, and/or planning to redeploy business modules as corporate modules.<br>4. EPA's internal vision for integration of information beyond the regulatory/ambient information realms is clarified. | **2001**<br>1. Exchange Network governance and interagency roles are established and designated people are in place.<br>2. Scope of the first phase of the Exchange Network is fully defined and operational for a limited subset of shared environmental business functions between EPA and a few prototype states.<br>3. States and EPA are actively engaged in defining subject matter area 'business model' neutral exchange formats, and retooling existing information exchanges towards the adoption of these formats.<br>4. State talent and motivation is capitalized on to create as many transaction sets for the Exchange Network as possible.<br>5. The path towards expansion of the Exchange Network beyond EPA and state environmental agencies is well understood. A clearer vision of the Exchange Network's second phase of development and use is established.<br>6. The Exchange Network is trusted, and all security concerns are reviewed, well understood and properly addressed.<br>7. EPA and States have determined how best to address the management of the Exchange Network. | **2001**<br>1. Initial prototype pilots have identified readiness factors (technical, policy, and organizational) for Exchange Network participants.<br>2. How best for EPA to assist its partners prepare for participation in the Exchange Network is clear and well understood. From these readiness factors an Action Plan for assistance to Exchange Network participants is fully defined.<br>3. A state/EPA Action Team is actively assisting states evaluate their readiness to participate in the Exchange Network |

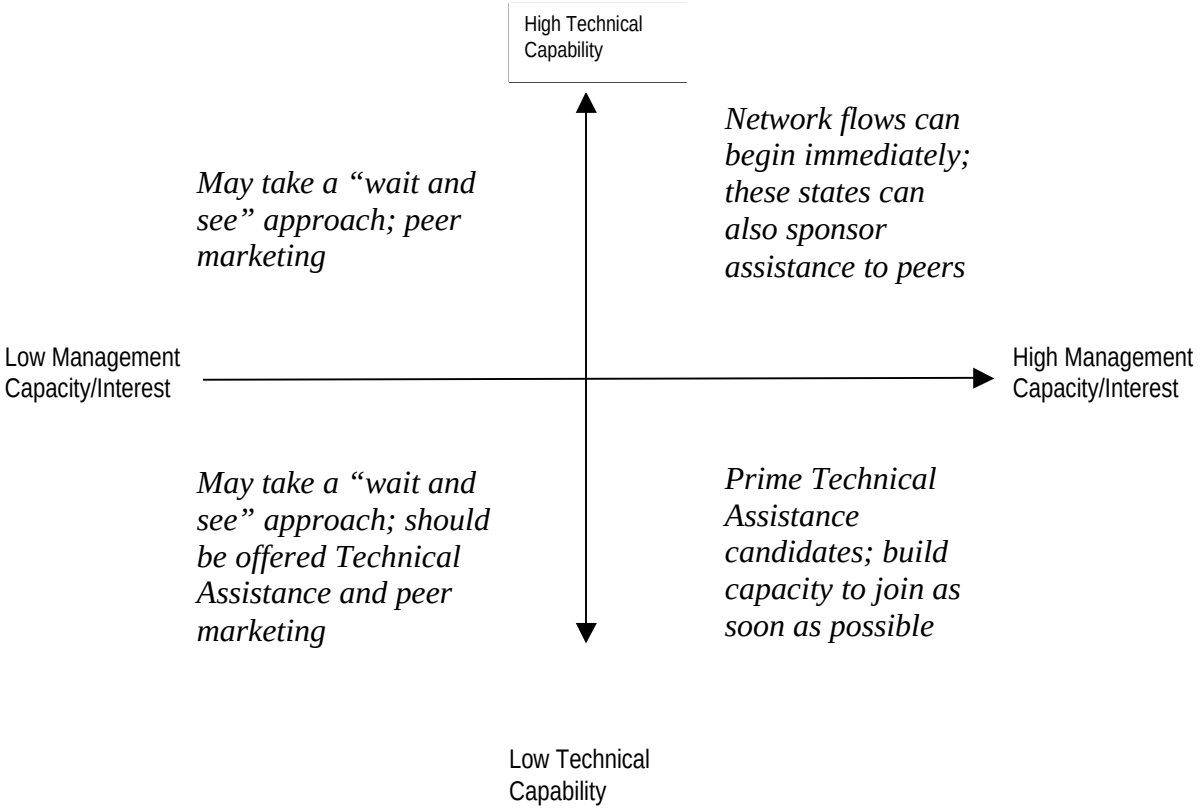
| EPA Internal Integration | Exchange Network Infrastructure | Partner Assistance |
|---|---|---|
| **2002**<br>1. The Enterprise Architecture continues to serve as a strategic framework upon which programmatic IT investment decisions are made, with appropriate revisions included.<br>2. Transition plans are in place and migration to the Enterprise Architecture is underway for EPA's major program systems and second tier systems.<br>3. An expanded set of foundation components (system of registries, business modules, and functions) serves as the authoritative source of key agency data and functions.<br>4. The third tier of foundation components is under development.<br>5. EPA Programs and Regions have migrated all major and some minor systems to utilize the core infrastructure (where appropriate) and have demonstrated benefits in terms of increased access and analytical capacity and increased efficiency in utilizing resources.<br>6. I-3's progress and the approach in its management plan are reviewed. | **2002**<br>1. Governance and stewardship of the Exchange Network are routine operations<br>2. Exchange Network is operational.<br>3. The Exchange Network has reached out beyond EPA and state environmental agencies and is operational with other parties.<br>4. States, EPA, and the new partners are continuing to retool existing information exchanges towards the adoption of agreed-upon formats. Work is underway in identifying new partners with whom information exchanges need transformation.<br>5. The Exchange Network is trusted, and all security concerns are reviewed, well understood and properly addressed | **2002**<br>1. A mechanism for assisting partners to assess their readiness to function as Exchange Network portals continues to operate.<br>2. A mechanism for technical assistance to trading partners to implement and secure their Exchange Network portals continues to operate.<br>3. An assistance mechanism for states/partners to participate in developing exchange formats is operational. |

## State Organizational Infrastructure

Each state environmental agency will need to assess its own information management status and level of readiness to join the Network. Three levels of overall technical and management readiness can be examined. (This concept of "readiness" is borrowed from the e-commerce network vocabulary.) Large firms (e.g., IBM or Intel) have begun to formally assess the readiness for e-commerce partnerships with their suppliers and distributors. An excellent technical overview of this process as it applies to e-commerce is included in the "RosettaNet" paper included in the Blueprint reference materials.

As indicated in the chart below, technical capacity can be thought of as the ability to build a node and the internal systems feeding that node. This is a relatively traditional software/Web development task. Management capacity and interest are different and more complex. They include the internal discipline and coordination to ensure that high quality data is available to the

node manager, and that TPAs covering that data can be negotiated and implemented. "Interest" is included in this category because the Network is voluntary and the first flows will require pro-active involvement on both sides.




This chart is useful because it depicts a wide range of possible starting points for any given state (and for EPA as an agency). State participants in the Blueprint Team span the spectrum of capabilities and interests identified in the table.

- At the highest level of overall readiness are some states (top right of chart) with robust technological and management data exchange capabilities. Several of these states have participated in Network pilots and other projects and could begin Network flows within months. These states are also in an excellent position to partner with EPA and use the IMWG's Knowledge Transfer Action Team to share their experience (and perhaps specific tools and approaches) with other states.

- States in the center of this chart enjoy some of the infrastructure needed, but need further development of some technological or programmatic components in order to join the Network; they are ideal targets for the Knowledge Transfer Action Team since they are almost ready to go.

- The Network design emphasizes open and flexible tools for partners to use; therefore, few states need find themselves in the lower right hand portion of the chart. With sufficient management commitment and some support, most states should be able to

participate. It may be appropriate for these states to initially rely on EPA for some portions of their technical infrastructure. This would be similar to a state's decision to use EPA's CDX as its electronic reporting infrastructure, while maintaining stewardship/ownership of that data.

- The few states on the left of this chart pose a different challenge. In the upper left (which is believed to be nearly empty) states have the technical capacity to participate in the Network (perhaps because they are already building portals of their own, and are familiar with XML technologies) but do not have the management capacity or interest. These states should be the targets of "marketing" outreach efforts. EPA will need to ensure that it is offering these states the ability to transition flows to the Network.

- Finally, states in the lower left corner should be offered technical assistance and "peer" outreach, especially to ensure that what appears to be a lack of management commitment is not actually concern that the technical threshold for participation is just too high to merit management investment.

- A special focus similar to the Knowledge Transfer Action Team's "Small States" working group may be an excellent way to document and share the experience of states that have rapidly moved up and into Network participation.

## IMWG Organizational Infrastructure

The IMWG is the core forum for state and EPA collective action. As such, the IMWG will play a crucial role in creating Network flows. The IMWG chartered this Blueprint development effort. However this Blueprint does not recommend that the workgroup *be* the Network, nor its administrator. Instead, this Blueprint proposes that the IMWG be the venue through which the institutional home and capacity for these functions be identified and launched. In addition, the workgroup is the only body that can provide some high-level coordination and support to ensure that its own sponsored activities are advancing the Network. This coordination and support is available through several means:

- Data Standards Council
  - Encourage DET developers to use the DCS
- Central Data Exchange Action Team
  - Forum for data exchange issues as they relate to EPA's CDX
  - Two Network flow projects (DMR and STORET) launched by the team
  - Security/E-commerce interoperability project
- PCS/IDEF Action Team
  - Encourage/support use of Network concepts as final design for IDEF is established
- Facility Action Team
  - Establish flows of facility data
  - Evaluate TPAs for facility data

In addition, the IMWG is the only organization positioned to support some of the broader intergovernmental commitments and expectations, such as the following:

- Commitment on behalf of enough participants to conduct business in the new manner to make the endeavor worthwhile.
- Commitment on behalf of participants to take on the implied data stewardship responsibilities[11].
- EPA commitment to retool existing state reporting relationships to accommodate Network principles.
- Commitment on behalf of participants to financially support the Network and work towards establishing self-sustainability for this function.
- EPA commitment to investments for accelerating the Network and support for the DET development process so that DETs are, or can be, available for those who want to use them.
- EPA commitment to maintaining multiple (old and new) business practices for receiving data from partners so that the Network is truly voluntary.

[11] Each participating agency, as a Network partner, in agreeing to host their information, assumes data management responsibility for their portion of the Network. Data quality, timeliness, error correction, metadata expectations, and standard operating procedures will all need to be developed, built into transaction set requirements, and incorporated into TPAs. (The degree of oversight and specificity would vary depending on nature and granularity of the exchange.)

# 11. Relationship of Network Components

The matrix on this and the following page describes the relationship of each of the Network components to the other components.

| | | Data Standards | Data Exchange Templates |
|---|---|---|---|
| | **Data Exchange Templates:** | – Data standards will be incorporated into Data Exchange Templates<br>– Cross-program data standards implemented in DETs will improve integration. | |
| | **Trading Partner Agreements:** | – Trading Partner Agreements will identify which Data Standards are being used. | – Trading Partner Agreements will identify which Data Exchange Templates are being used. |
| | **Technical Infrastructure:** | – The Technical Infrastructure (e.g. XML schema) will validate that a Data Standard is being used.<br>– The Technical Infrastructure will provide easy/open access to all official Data Standards. | – The Technical Infrastructure will validate that a given transmission is compliant with its Data Exchange Template. |
| **Organizational Infrastructure** | **Network Governance:** | – Network Governance for Data Standards will be through the Environmental Data Standards Council. | – Coordination/governance of Data Exchange Template development is conducted, especially for the traditional state/EPA data flows. |
| | **EPA:** | – EPA will develop policy enforcing the use of Data Standards in all internal information management activities. | – National programs identify priority areas for Data Exchange Templates development between states and EPA |
| | **State Environmental Agencies:** | – State Environmental Agencies will develop policy around the use of Data Standards developed by the Environmental Data Standards Council in all information management activities. | – National programs identify priority areas for Data Exchange Templates development between states and EPA<br>– State-to-state flows use Data Exchange Templates. |
| | **State/EPA Information Management Workgroup:** | – The State/EPA Information Management Workgroup will continue to provide support to the Environmental Data Standards Council. | – The State/EPA Information Management Workgroup will provide guidance for how to develop Data Exchange Templates. |

| | | Trading Partner Agreements | Technical Infrastructure |
|---|---|---|---|
| | **Data Exchange Templates:** | | |
| | **Trading Partner Agreements:** | | |
| | **Technical Infrastructure:** | – The Technical Infrastructure will include a neutral repository where Trading Partner Agreements will be posted. | |
| **Organizational Infrastructure** | **Network Governance:** | – Governance is needed for the Trading Partner Agreement format and development mechanism.<br>– Oversight of Network expansion to additional data partners will occur. | – Policies will be established to define general security processes used on the Network. |
| | **EPA:** | – Trading Partner Agreements document official flows for regulatory reporting requirements from states to EPA program offices.<br>– Oversee Regional role in the governance of Trading Partner Agreements. | – The availability of EPA funding will affect the ability to assist states in developing technical capacity. |
| | **State Environmental Agencies:** | – Trading Partner Agreements document official flows for regulatory reporting requirements from states to EPA program offices.<br>– States will coordinate the management of Trading Partner Agreements with their EPA Region. | – Agencies will coordinate/leverage state technical investments via Knowledge Transfer.<br>– States will have some ability to influence EPA's technical decisions/investments. |
| | **State/EPA Information Management Workgroup:** | – The State/EPA Information Management Workgroup is responsible for oversight of and coordination of the Trading Partner Agreement framework for state/EPA data flows.<br>– The State/EPA Information Management Workgroup is responsible for coordination of Network expansion to additional data partners. | – The State/EPA Information Management Workgroup comments on technical standards that influence technical infrastructure. |

# 12. Recommendations to the Workgroup (October 2000)

Based on the discussion and analysis documented above, the Blueprint Team makes the following recommendations to the IMWG:

- The IMWG should approve the Blueprint.
- The Network Blueprint Team should stay intact to develop a specific proposal on Network administration that includes financing options.
- The IMWG should identify its next steps in advancing the Network, including a plan for outreach.
- The process of using the Network components to build Network flows should begin immediately.

Note: The Workgroup formally endorsed this report and the above recommendations at its October 2000 meeting.

# 13. Network Example

## Multi-State Watershed Project

Appendix F presents an example of a voluntary agreement among trading partners for the purpose of exchanging data. Not every Network TPA would follow this format, just as many other types of state/EPA operating agreements look quite different from case to case. The particular circumstances of the parties involved and the data being exchanged will influence which elements are included in the agreement and how these issues are described.

# 14. Addendum: Network Administration Report to IMWG

**Note: These recommendations and Initial Implementation Proposal On Network Administration were prepared for the IMWG Meeting on February 6 and 7, 2001. The IMWG approved action on all recommendations and endorsed incorporation of this report in the Blueprint.**

## Introduction

This report provides background and recommendations of the Network Blueprint Team under its October 2000 charge from the State/EPA Information Management Workgroup (IMWG) to further develop the Network administration functions described in the Blueprint for an Environmental Information Exchange Network. *The Team requests the IMWG to pay particular attention to the recommendations and requested decisions (summarized below).*

The report is organized as follows:

Section 1 contains a summary of the recommendations and requested decisions.
Section 2 contains a more detailed description of each recommendation.
Section 3 describes the Network prioritization effort.
Section 4 defines specific Network administration functions.
Section 5 discusses options, including a potential role of a third party in sourcing these functions.
Section 6 provides an implementation framework and schedule for the Network.

## 1. Summary Recommendations of the Blueprint Team on Network Administration

The table below outlines the recommendations of the Network Blueprint Team for current IMWG action. Detailed descriptions of these recommendations are included in Section 2 of this report.

| Recommendation | February Action Requested and Approved |
|---|---|
| 1. Charter Interim Network Steering Group | Approve charter and authorize immediate formation of Interim Network Steering Group |
| 2. Establish Network Registry Test Bed | Approve development of test-bed registry under the guidance of the Interim Network Steering Group to be operational by March/April 2001 |
| 3. Propose DET Development/ Harmonization Approach | a) Charge the Facility Action Team to finalize a facility DET and assess its applicability to facility data in other flows<br>b) Charter a research effort to produce recommendations on Network DET harmonization by Summer 2001 |
| 4. Investigate Third Party and Other Options for Sourcing Network Administration | a) Approve continuing research effort on the role of third parties and other sourcing options<br>b) Approve development of a semi-formal request for information from candidate institutions/forums in Spring 2001. The informal request would be presented before release to the IMWG at its next meeting. |

| Recommendation | February Action Requested and Approved |
|---|---|
| 5. Develop Implementation Plan | a) Review and discuss the high-level framework and major milestones presented here<br>b) Authorize continued refinement of the plan by the Interim Network Steering Group for presentation to the IMWG at its next meeting |
| 6. Coordinate Environmental Data Standards Council (EDSC) and Network Steering Group | Charge the Interim Network Steering Group and EDSC to jointly consider coordination issues and provide recommendations to the IMWG at its next meeting |

# 2. Description of these Recommendations

1. **Charter Interim Network Steering Group:** This report identifies a critical near-term role for a Network Steering Group, especially in oversight and coordination of the near-term activities proposed here. The Network Blueprint Team recommended that the Network Steering Group function be fulfilled in the short term by an Interim Network Steering Group. This group would be composed primarily of the members of the existing Network Blueprint Team, and would focus on near-term Network implementation and plan for a December 2001 sunset period. A draft charter is attached for IMWG's review that defines the mission, objectives, scope, membership and schedule for an Interim Network Steering Group. The Blueprint Team requests this rechartering to formalize IMWG approval of its transition to this new more operational role.

   February Action Requested and Approved: Approve charter and authorize formation of Interim Network Steering Group. This draft charter contains a sunset provision and will be superceded by whatever Steering Group is established once the Network Administration function is formalized and sourced by the IMWG.

2. **Establish Network Registry Test Bed:** In the simplest sense, the benefits of XML will only be achieved if a significant number of organizations are using the same XML documents. Therefore, these XML documents must be available for partners to *discover* and *retrieve*. A registry/repository is a mechanism used to *discover* and *retrieve* documents, templates, or software (i.e., objects and resources) over the Internet. A *registry* is the mechanism used to *discover* the object. The registry provides information about the object, including its location. A *repository* is where the object resides. A user *retrieves* an object from a repository. The Network registry would be the first step to supporting harmonization and integration of DETs. Similar registries are a critical core component of comparable private and mixed sector networks. Significant research by the Network Blueprint Team and the EPA XML TAG into the requirements for a test-bed registry indicates that the work of other organizations (e.g., NIST and OASIS) can be used to develop a prototype registry. This would help ensure conformance and interoperability with emerging registry standards and provide a jump-start to the overall effort. A small team of expert state and EPA staff would oversee mounting and initial operation of this registry. This expert group would report to the Interim Network Steering Group once that group is chartered.

   February Action Requested and Approved: Approve development of test-bed registry under the guidance of the Interim Network Steering Group. The Network Blueprint Team seeks to have the registry operational in the March/April 2001 timeframe.

3. **Propose DET Development/Harmonization Approach:** DET development appears to be accelerating. As it does so, DETs will proliferate. Harmonization of these DETs towards a broadly compatible and consistent framework is a long - term objective of the Network. Because standards and methodologies for harmonization are still very much evolving, the Blueprint Team recommends two parallel efforts for the near term: a) charge the Facility Action Team (in coordination with others) to develop a first-generation DET for facility data, and to use that experience in guiding the harmonization of facility data in other DETs under development; and b) commission development of a short white paper that recommends medium-term investments in Network DET harmonization. These recommendations will likely include additional requirements for registry operation, development of reference models, and specific areas and proposals for coordination with the following: ongoing DET development of the IDEF and CDX Action Teams, the Environmental Data Standards Council and its approved standards, the EPA XML TAG and others. These recommendations will be vetted by the Interim Network Steering Group and then presented to the IMWG for final approval in early Summer 2001.

   February Action Requested and Approved: a) Charge the Facility Action Team to finalize a facility DET and assess its applicability to facility data in other flows; and b) Charter a research effort to produce recommendations on Network DET harmonization by Summer 2001.

4. **Investigate Third Party and Other Options for Sourcing Network Administration Functions:** This report proposes specific roles for the Interim Network Steering Group; additional refinement of these and other roles will be developed in the Implementation Plan (below). As discussed below, the Network Blueprint Team has also identified specific functions that may benefit from the use of a third party. (Please see Section 5 for a detailed list of these functions.) Possible candidate third parties have also been identified. The Team recognizes that third parties or others may be able to perform certain Network administration functions within a relatively short timeframe; therefore, investigations into the capabilities of candidate organizations and careful consideration of the resulting options should begin as soon as possible, including options other than third party. The Team requests authorization for the Interim Network Steering Group to recommend desired capabilities and functions of a third party, for presentation to the IMWG at its next meeting. The Team proposes to use these recommendations, as approved by the IMWG, to solicit an informal "request for information" to identify capable third parties and assess their capacity to fulfill the identified functions as compared to other options.

   February Action Requested and Approved: a) Approve continuing research effort on the role of third parties; b) Approve development of a semi-formal request for information from candidate institutions/forums in Spring 2001. The informal request would be presented before release to the IMWG at its next meeting.

5. **Develop Implementation Plan:** This report provides a draft high-level framework and milestones for near- and mid-term Network implementation (see Section 6 for this implementation plan). The current Network Blueprint Team proposes to continue refining a more detailed implementation plan under the Interim Network Steering Group

and to present the proposed plan at the IMWG's next meeting. This refinement would include merging the Network priorities for 2001 and identified needs for 2002 with the plan. This may require adjustment of the milestones identified below.

February Action Requested and Approved: a) Review and discuss the high-level framework and major milestones presented here; b) Authorize continued refinement of the plan by the Interim Network Steering Group for presentation to the IMWG at its next meeting.

6. **EDSC and Network Steering Group Coordination**: Many of the support functions identified as candidates for third party involvement could also apply to the EDSC. In addition, close coordination of the EDSC and Steering Group will be important in ensuring standards support harmonization as identified above. A meeting between EDSC and Network Blueprint Team leads is planned to discuss coordination of support functions such as staffing, contractors and travel.

   February Action Requested and Approved: Charge the Interim Network Steering Group and EDSC to jointly consider these issues and provide recommendations to the IMWG at its next meeting.

## 3. Network Project Prioritization Effort

Responding to their charge at the December 21, 2000 IMWG conference call, the state and EPA co-chairs of the Action Teams and Data Standards Council engaged in a process of identifying and prioritizing near-term projects to support implementation of the National Environmental Information Exchange Network.

The co-chairs group met via conference call three times; on January 3, 5 and 8. At the first meeting, the group agreed upon a list of ongoing and planned projects that will support Network implementation. After the first meeting, members of the group used established criteria to rank each project as high, medium, or low priority. Thirteen projects were forwarded to the EPA Quality Information Council (QIC) as high priority projects, along with several projects ranked high with issues or questions.

The six recommendations in Sections 1 and 2 of this report link directly with the priorities list generated by the co-chairs. Although not included as a specific high priority project, construction and operation of the Network Registry is recognized by the Network Blueprint Team as an effort that supports other high priority areas (e.g., DET Development and Harmonization) and should proceed as was planned before the prioritization exercise.

## 4. Network Administration Functions Defined

As defined in the Network Blueprint, Network administration includes those infrastructure functions that support Network operation and are outside the core environmental management functions of the participants. Network administration will not include inherently governmental functions: it will simply support the flow of data through the Network. The Network Blueprint Team followed several basic principles in creating recommendations in this area: 1) the administration functions will be kept to the absolute minimum needed to start the Network; 2)

the administration functions will be expanded to provide more functionality as it becomes necessary and credible to do so and 3) Network administration will have some degree of autonomy from individual members but will service the trading partners and be transparently accountable to them.

## Team Approach

The Team used the following approach in development of these recommendations:

- Re-established and reclarified the validity of the Network administration concept as an essential and credible function for Network operation.
- Using several working sessions, meetings and reviews by outside experts, sought to identify further required Network administration functions. This effort largely confirmed the high-level areas identified in the Blueprint.
- Scrutinized these functions, in some cases breaking them into smaller, discrete areas (e.g., several kinds of required registries or official lists). These functions were then regrouped into the more manageable list provided below. Consideration of the "registry" and "steering group" functions received the majority of the T attention here.
- For each of the major functional areas identified, deliberated the minimal structures and policies needed to begin flows. In addition, the Team evaluated the costs and benefits of having semi-autonomous (defined below) third parties conduct some aspects of some of these functions on behalf of the states and EPA.
- Developed a rough implementation proposal for these functions.

The team's analysis and deliberations produced a greater level of supporting detail than is presented here. With approval by the IMWG (see recommendations) this information will be further refined, reviewed and incorporated into the implementation report prepared by the Interim Network Steering Group for the IMWG February 2001 meeting.

## Major Network Administration Functions Identified

The Team identified eight core functions of Network administration, many of which overlap and are inter-related. The first four functions focus on the *information:* its structure and consistency are especially inter-related and synergistic. For example, the establishment of a registry of Document Exchange Templates (DETs) is the first essential step in coordinating development efforts, providing a starting point for new DET development, and beginning the work of harmonization. The last four functions focus on the *administration* of Network members and support of their capabilities. All of these functions are envisioned as being coordinated and overseen by the Network Steering function (i.e., Steering Group) described below.

1. **Administration of Registry for Network DETs** – A central registry and repository where DETs are housed on a voluntary basis for participant's reference and use. Initially, this may be a simple list of draft DETs, but can expand to a searchable tool to locate DETs for specific environmental business areas with links to other relevant registries. This registration may be a requirement of a Trading Partner Agreement (TPA). This registry may also eventually link to other registries, including those of data elements and metadata such as the Electronic Data Registry (EDR).

2. **Administration of a Repository for other Network Administrative References** – A reference library function for use by participants looking for example TPAs or information about the Network, such as the following:
   - DETs known to be under development
   - TPA templates and executed TPAs
   - Network membership roster
   - Network policy and technology guidance documents

3. **Process and Technical Support for DET Development –** Support and guidance to participants who want assistance in creating DETs. This may include direct guidance for using XML (eXtensible Markup Language), incorporating data standards, and composing a DET. It may also include the development of DET templates.

4. **Harmonization and Use of Data Standards In DETs** – Active encouragement of harmonization of DETs through the use of reference models, data standards, standardized DETs, and other approaches as appropriate.

5. **Technical Infrastructure Development**
   - Technical assistance, such as developing and distributing readiness assessments and quick start kits for becoming a Network node.
   - General security policies and tools, for example on levels of security available on the Network and options for technology used. More specific security measures between parties may be detailed in individual TPAs.

6. **Communications, Outreach, and Inreach to State and EPA Staff; Liaison with External Groups** – As directed by the Steering Group, support for communications, outreach and inreach to stakeholders.

7. **Network Membership** – Targeting of potential participants, dissemination of membership information, and liaison and coordination with external parties/consortia, especially those in related subject areas (e.g., Land XML, Chemical, EHS).

8. **Network Steering** –Venue for development of Network policies and practices, as distinct from those of an individual agency. The steering function also coordinates the activities of Network sub-teams and may provide direct oversight of contractors or third parties acting on the Network's behalf. The team recommends that this function be fulfilled by an Interim Network Steering Group rechartered from the Network Blueprint Team.

# 5. Third Party and other Options for Sourcing Network Administration

Since receiving the charge to develop Network administration recommendations, the Team has focused on specific administrative functions and their execution. Before reviewing possible roles

and responsibilities, it may be helpful to distinguish the following modes of participation in the Network and its development:

- *States and EPA staff acting independently on behalf of their own programs and agencies.* Example: state and EPA staff negotiating a TPA for a specific flow.

- *States and EPA acting as part of an IMWG/Network sponsored group*. Example: state and EPA staff collaborating on Network security protocols.

- *Contractors supporting states or EPA in either of these capacities*. At the direction of their state or EPA clients, contractors providing technical analysis, meeting support, facilitation, XML expertise, and other services.

These traditional roles have supported the IMWG and groups like the Environmental Data Standards Council to date. The Team expects that these roles will continue. As indicated in the recommendations above, the Team anticipates that the Interim Network Steering Group will perform or coordinate many of the eight functions identified above. Many of the private/mixed sector networks benchmarked by the Team make use of actors termed "third party" in the sense that they support trading partners in a semi-autonomous role. Based on initial research, the e-commerce networks reviewed by the Team clearly indicate a role for such parties in establishing the procedures and supporting the administrative tools used by trading partners. For example, the RosettaNet trading network consists of bilateral or multilateral agreements between partners to conduct business using a set of standards. The RosettaNet.org organization is the third party: it fulfills many of the functions discussed below but does not become involved in the business or disputes of trading partners. The broader ebXML initiative uses a similar ebXML.org organization hosted in the United States by OASIS.

The Team found it most useful to consider the option of engaging a third party on behalf of the Network in the context of specific Network functions only. At its plenary meeting in Annapolis, the Team candidly reviewed its detailed list of functions and considered if and how a third party acting on behalf of the Network could provide a unique capability. This detailed discussion and the research conducted so far yielded the following general principles:

- Third parties may provide a unique capability in establishing the credibility of the Network. EPA and states envision this Network as a means to transform their information relationship and eventually expand it to include many other stakeholders. The Network administration must be credibly independent of any one member and accountable to its joint/independent Steering Group. For example:

  - If Network administration were perceived as "EPA" owned, or as the prerogative of some individual state or state organization, members could easily interpret actions or glitches in administration to be politically motivated. This would drag the Network backwards into previous unproductive conflicts.

- Network administration functions must provide a neutral forum where states, EPA, and as approved, other stakeholders, can be involved in data-centric discussions independent from the complex bilateral regulatory obligations concerning this data. This forum also provides the opportunity to harness peer learning and peer pressure towards standardization.

- The rate at which XML vocabulary consortia are being created in similar or overlapping areas (e.g., LandXML, Biology-XML, Environmental Health and Safety, Petroleum Industry XML) is likely to continue or accelerate. In addition, the environmental software vendor community, as well as the regulated community, are rapidly embracing XML as the interchange technology of choice. These groups are likely to be planning consortia or other associations to advance these interests.

### Assumptions and Risks

- As noted in the Blueprint, neither the Network itself, nor any third party engaged to support it, can perform any inherently governmental functions. These are and remain the purview of the independent member agencies. Third parties can neither get involved nor risk getting involved in jurisdictional or programmatic negotiations between members, especially states and EPA.

- Third parties participating in Network administration must be absolutely neutral, vendor-independent and apolitical. Third parties can bring technical and standards-related expertise and perspectives to support these aspects of the Network, but they should not have their own environmental policy or programmatic agenda. Groups like ANSI or OASIS fulfill these criteria because they are standards focused.

A more detailed table of specific functions and possible roles for a third party can be found in the table at the end of this report.

## 6. A Draft Implementation Framework for the Network

The Network Blueprint Team recommended that the Network Steering Group function be fulfilled in the short term by an Interim Network Steering Group. The attached draft charter outlines the scope and charge of this group.

The Network Blueprint Team has developed a high-level implementation framework to be carried out initially by the Interim Network Steering Group for 2001. The Team recommends that this framework be merged with results of the prioritization effort, revised, and presented to the IMWG at its next meeting. The Team recommends that states and EPA begin actively preparing for major Network infrastructure decisions and investments to be made towards the end of 2001. By the first half of 2002, the Team expects the Network to be able to support scores of official flows across many program areas with a stable, organizationally secure infrastructure. This implementation framework can be separated into the following timeframes:

## February 2001 – March 2001

Planning themes: *Implement and begin using test registry; confirm alignment of current flow projects; Network Blueprint Team continues as Interim Network Steering Group.*

Network priorities: Network Steering and Implementation Plan Development, DET Development, Air emissions, IDEF and FRS Pilot Flows

Milestones:

- Test registry operational
- Interim Steering Group chartered and in operation
- Facility Team working to harmonize facility data DETs
- ECOS Annual Meeting (2/2001), preliminary Network flows demonstrated (PCS, Facility, Air)
- IMWG Meeting (2/2001) Decisions on next implementations steps

## April 2001 – August 2001

Planning themes: *Learn by doing; achieve first "official" Network flow and prepare for semi-formal request for information for sourcing. Prepare for next round of major implementation decisions.*

Network Priorities: Pilot flows, DET development, Network Steering and Implementation Plan Development

Milestones:

- First "official" Network flow established (i.e., flow that satisfies a formal obligation and replaces an existing flow)
- Test DET registry in use for evaluation and first "official" flow
- Prototype administrative registries operational
- Structure and approach for DET harmonization underway
- Preparation and issuance of semi-formal request for information on capabilities to support Network administration completed

## September 2001 – December 2001

Planning themes: *Assess experience gained from first flows, use of registry and information request; formalize Network Steering Group; prepare Phase II implementation recommendations for IMWG.*

Network Priorities: Security Assessment, Network Outreach, DET Development, Network Steering and Implementation Plan Development

Milestones:

- Independent security assessment and protocol revision completed
- Strategic engagement of external stakeholders underway
- Initial guidance on DET harmonization prepared
- Results from semi-formal request for information received and assessed
- Phase II Plan provided to IMWG
- IMWG decisions made on next implementations steps including:
  - Next-generation registry and registry hosting
  - Charter for permanent Network Steering Group

## Resources Required in this Implementation Framework

The IMWG has accomplished much with limited resources; much of this progress has depended on the contributions of staff (EPA and state) time on an informal or semi-formal basis. The next stage of development and implementation of the Network, however, will require additional resources for dedicated staffing and administration. The Network prioritization effort clearly identified the projects in greatest need of resources. This support is needed to ensure that the Network is able to credibly involve new states and programs and convince them to invest their efforts in flowing their data using the Network. These funding needs will likely grow as agencies move towards full implementation of their nodes once the proof-of-concept period has ended; it is also at this point that agencies will realize savings or costs avoided as these nodes replace current flows or are incorporated into ongoing system redevelopment.

## Appendix: Potential Third Party Functions

At its Annapolis meeting in December 2000, the Network Blueprint Team agreed that administration of certain nongovernmental Network functions by a third party could enhance the Network's success and credibility. Such a third party could act as a neutral forum for administrative and technical functions, but would not become involved in the policy or programmatic disputes of Network members. The third party may be a non-profit institution providing a suite of services to Network members, a host organization Network members may affiliate with or join that specializes in the information exchange field and can provide a base of support and knowledge to Network administration, several different organizations supporting specific administrative tools used by trading partners or other variations. Whatever its form, the third party would be accountable to the IMWG through some mechanism (e.g., contracts or cooperative agreements with EPA or another agency).

The table below lists several Network administration functions in which a third party could provide additional or unique value. Related IMWG/Steering Group roles are also listed. The Team recognizes that some functions, such as administration of a Network Registry and support of Network steering, represent more near-term opportunities for third party involvement, while other functions would be researched and developed through a Request for Information later in the year.

This list does not include functions similar to work routinely performed by agency staff or contractors (such as facilitation, research analysis or consultation). It is expected that staff or contractors would continue to provide this support even though some of these functions may also be provided by a third party at some point.

| Network Administration Function | Related IMWG/Steering Group Roles | Recommended Roles/ Advantages for Third Party Use |
|---|---|---|
| **Administration of a Registry for Network DETs**<br><br>**Administration of a Repository for other Network Administrative References** | ➢ *Establish registry policy*<br><br>➢ *Establish registry functional requirements* | ➢ *Build, operate maintain registry infrastructure*<br>– Provide host site for on-line registry/repository<br>– Design and build registry<br>– Receive DETs or other documents from Network members and make available to other members through registry site<br>– Establish and maintain security of registry<br>– Administer technical maintenance, troubleshooting of registry<br>– Maintain consistency with broader registry standards<br><br>➢ *Separation of registry operation from political and programmatic issues* |
| **Provide Process and Technical Support for DET development**<br><br>**Encourage Use of Data Standards in DETs** | ➢ *Prioritize support for DET development and harmonization* | ➢ *Provide electronic and organizational infrastructure (e.g., mailing lists, archives, and administration)*<br><br>➢ *Provide an objective, external and comprehensive view of DETs both in the Network and in other public and private arenas* |
| **Support Technical Infrastructure Development** | ➢ *Prioritize and sponsor technical infrastructure development* | ➢ *Provide an independent assessment of security and other infrastructure to increase confidence and credibility*<br>– Analyze registry operations, data flows, security operations and policies, and other technical elements<br>– Provide assessment of technical infrastructure and recommendations for improving technical operations of the Network |
| **Conduct Communications, Outreach and Inreach to State and EPA Staff; Liaison with External Groups** | ➢ *Establish outreach plan*<br><br>➢ *Establish strategy and timing for engagement of external parties* | ➢ *Advise Steering Group on potential participants*<br><br>➢ *Leverage pre-existing membership governance structures* |

| Network Administration Function | Related IMWG/Steering Group Roles | Recommended Roles/ Advantages for Third Party Use |
|---|---|---|
| **Promote Network Membership** | | ➢ *Provide global perspective on potential public and private members*<br><br>➢ *Use pre-existing connections to associations and consortia as means of finding relevant parties*<br><br>➢ *Assist new trading partners in setting up Network nodes* |
| **Network Steering** | ➢ *Establish basic Network policies*<br><br>➢ *Provide oversight of Network support and third parties* | ➢ *Provide neutral forum for Network Steering deliberations*<br>– Provide administrative and logistical support for meetings<br>– Provide neutral forum for policy development discussions<br><br>➢ *Provide technical advice on registry, data standards, DET development, XML, security, or other issues as necessary* |

# References


## State/EPA Information Management Workgroup

*Shared Expectations of the State/EPA Information Management Workgroup for a National Environmental Information Exchange Network*. Working Version, June 12, 2000.

## EPA White Papers

*The Exchange Network.* Draft August 1, 2000.

*Shared Network Governance and Stewardship of Data and the Exchange of Data*. Draft June 21, 2000.

## Industry White Papers

Sachs, et al., *Executable Trading-Partner Agreements in Electronic Commerce*. IBM T.J. Watson Research Center, 2000.

O'Sullivan, Patricia J. and Don S. Whitecar, *Implementing an Industry e-Business Initiative: Getting to RosettaNet*, Intel Technology Journal Q1, 2000.

## Appendix A:
## NETWORK FAQS

1. What kind of data might be on the Network?

Any member data that incorporates the components [see “what is a component” below] might be on the Network. At first, much of this data will likely be traditional information typically reported to EPA by state agencies or data that states share among themselves. The Network Blueprint Team expects (and hopes) that this will quickly expand to include a wide and diverse scope of data, ranging from regulatory to ambient, that incorporates the components.

2. Is all Network data in data exchange templates?

Yes, all the data mounted by a Network member is formatted (and therefore available) in a specified and registered format (Data Exchange Template). This is not a severe restriction because any member is free to propose and then use a DET for its own purposes, as long as their trading partner has agreed to that DET. Or, in the case of a single-party network declaration, as long as that DET is registered, the member is free to use it. The Blueprint Team expects EPA will establish a small set of nationally consistent DETs for its key data needs. EPA will accept these (and only these) DETs in fulfillment of specific reporting obligations.

3. Do Network flows replace existing uploads of data from states to EPA?

The aspiration is that Network flows would replace existing uploads of data from states to EPA. Using negotiated standards, DETs and TPAs, Network flows would be “official” and would therefore replace existing state to EPA programmatic reporting. This is why the blueprint details the levels of security and TPA assurances expected.

4. Who authorizes the Network?

The network will be created through a web of Trading Partner Agreements (TPAs) and core infrastructure components between its members. The Network Blueprint Team expects EPA and states to authorize individual flows that are designated in TPAs as "official" data flows and in doing so “authorize” the Network. The Blueprint Team also anticipates that EPA and states would officially express their interest, desire and investments in the development of the Network. States and EPA, acting through the IMWG, may also authorize a third party to act on their behalf as administrator of the Network.

5. Would it be appropriate for a state and EPA to link a Network flow and a grant?

Yes, it would be appropriate for a state and EPA to link a Network flow and a grant, <u>if</u> that is what is agreed to by EPA and that state. The Network Blueprint Team

believes such arrangements offer tremendous leverage to both states and EPA. EPA is able to fund development by states of stable, reliable environmental data services that benefit the nation. States could receive support from EPA to provide EPA needed data, but, by doing so through the Network, provide that data to all of its stakeholders.

6. Does the Network Blueprint involve standards for display tools or public access?

No, these issues are not described in the Network Blueprint. The Network describes a web of information nodes, some of which are restricted -- it does not describe, nor does it constrain anybody from displaying, any data to which they have authorized access in any way, using any tool. With the exception of uses or displays that violate TPAs or stewardship principles, the Network Blueprint Team expects trading partners to use a wide variety of display approaches for their own and others' data.

7. How will expansion of the Network be addressed?

Incrementally - beginning with states and EPA. It is anticipated that participants will learn lessons and establish sufficient infrastructure to enable others to join the Network as their interest and capacity allow. Much of the design of the Network is based on the need for it to scale up. The vision for the Network is of a structure that eventually evolves to include everything, such as data from volunteer monitoring efforts, regulated entities and local governments.

8. Won't the Network be prohibitively expensive?

The creation of quality data sources is always expensive. These costs are borne already but, the data created by many of our efforts are inaccessible. Much additional EPA and state investment seeks to make this data available – the proposed Network is an extremely cost effective way to do this. By leveraging existing, open, vendor-neutral (and in some cases public domain) tools, these costs can be shared and minimized. The Network Team anticipates that development of a state agency "node" will cost no more than the development of a typical program system (e.g., hazardous waste), but could serve the entire agency. EPA has already committed itself to development of the Central Data Exchange facility which will constitute a significant component of its Network Node.

9. Are e-commerce tools like XML and TPA's really applicable or practical for environmental agencies? We are government agencies not amazon.com or kozmo.com.

The Blueprint Team believes the tools and technologies being developed and rapidly embraced by the private sector can be applied to the business of environmental agencies. Much of this technology enables the rapid, secure and cheap formatting and transmission of data between entities, leveraged through tools like XML and e-commerce servers. The Blueprint analysis suggests that in most cases,

states and EPA can apply simplified (leading edge, but not cutting edge) versions of the more mature technologies directly to their data. The market place is developing tools to securely clear thousands of e-commerce transactions an hour between thousands of trading partners. The Network Blueprint Team expects to be able to use simple versions of these tools and approaches to build the Network between the 50 states, EPA and eventually others.

10. What will be the role of EPA Regions in the Network?

The role of EPA Regions is discussed at length in the Blueprint document. Because the Network relies on TPAs executed between partners, the Blueprint Team envisions that EPA Regions would assume this role. In addition, EPA Regions would then use these TPAs as the basis for monitoring the quality and availability of these data flows. In addition, EPA Regions may facilitate the negotiation of multi-state/party, geographically-specific TPAs and flows.

11. What will be the role of the Environmental Data Standards Council in the Network?

The data transmitted in Network flows will follow the standards developed by the Environmental Data Standards Council where such standards exist. As the Council offers additional data standards to the environmental community, Network participants will incorporate these standards into Network flows.

12. How can the IMWG advance implementation of the Network components?

The IMWG can strongly encourage that new data exchange projects it sponsors are consistent with and actively advance Network components. These projects may involve development of data exchange templates, data flows, or trading partner agreements that can build on the foundations of the Network Blueprint. The IMWG can also facilitate pilot testing that will help shape and refine the Network components. Specific recommendations for IMWG consideration are listed in the Blueprint.

**Appendix B:**
**Working Version**
Version 1.8.3 (6/12/2000)

# Shared Expectations of the State/EPA Information Management Workgroup

# for a National Environmental Information Exchange Network

## Part I

In 1998 States and EPA committed themselves to a partnership to build locally and nationally accessible, cohesive and coherent environmental information systems. This commitment was codified in the State/EPA Information Management Workgroup Vision and Operating Principles. Now with two years of joint experience, States and EPA have developed a more specific vision for how this partnership could be realized in the form of a national environmental information exchange network (Exchange Network). We expect this Exchange Network to revolutionize our management of environmental information. Over the next three-to-five years, the Exchange Network will increase our efficiency, improve the quality of our environmental data, provide our agencies and the public ready access to this data and increase their ability to employ this information to protect public health and the environment. This Exchange Network will be standards-based, highly interconnected, dynamic, flexible and secure, operating with broad-based voluntary participation of the individual States and EPA.

The Exchange Network=s design and operation incorporates the following principles:

- $ An agency is, by mutual consent between a State and EPA, explicitly recognized as the steward for specific environmental information that will become part of the Exchange Network.

- $ The steward agency manages its data, provides access to that data via the Exchange Network and is accountable for the data's quality and availability. For each set of information, stewards will also maintain and make available a standard set of descriptive information which will document the data=s quality, currency and context.

- $ States and EPA offices, whose use of stewarded data necessitates the maintenance of local copies, are responsible and accountable for ensuring the integrity and currency of those copies.

- $ The Exchange Network employs an agreed upon set of common data *exchange* standards and Internet protocols; it does not dictate or constrain internal agency systems, software and other tools.

The States and EPA expect this Network to replace and continuously refine many existing data flows. As it grows, the Exchange Network will allow participants to quickly and easily access and integrate high-quality data that they or other participants have provided. Members will use the network in the way that best meets their individual business needs and that supports improved environmental decision-making.

## Part II Operational Propositions

This vision represents a major but timely change in current practice and direction. We have not and should not yet define either the technical or organizational details of its operation but advance the following propositions about the operation of the Exchange Network:

- $ EPA and States ultimately envision a broad and diverse membership, linking local, state, Federal and Tribal agencies. We intend to begin the Exchange Network between States and EPA and to expand it as fast as our experience and the interest of others allows.

- $ We also know that our current data and information flows are not always sufficient for our individual and collective missions. While we intend to begin building and learning on the basis of our current data flows and obligations, we intend to use this experience to help us jointly identify, collect and exchange relevant information for ourselves and the public.

- $ This Exchange Network vision will be realized through three areas of State/EPA joint commitments:
    - o to harness the technologies of the Internet by making a small but critical set of technology decisions together;
    - o to develop and negotiate, through the Workgroup, the programmatic and operational procedures needed to begin the Exchange Network; and
    - o to strategically assess, invest in and monitor the technical and programmatic capacity of all States and all parts of EPA to use the Exchange Network; this strategy will acknowledge the broad diversity of current and future state approaches to information management and States= current and future actual uses of EPA systems for State management purposes.

- $ Many State and EPA members need to use information stewarded by others. We envision that, ultimately, technology will allow the real time, instantaneous manipulation of distributed Exchange Network data without the need for any party to maintain a local Awarehouse@. But this technology is not yet practical for all of the network participants and data sources. In the interim, many users will need to create and maintain (consistent with the responsibilities above) local copies of data stewarded by others. The operating principles of this Exchange Network are intended to reduce both member=s costs and to improve the integrity of these local versions by explicitly acknowledging the system of record, the stewardship of the data, and the quality and change authority responsibilities established through the Exchange Network operating procedures.

The shared goal of creating and using this Network provides renewed focus and direction for the collective work of the Workgroup. We intend to use the Exchange Network as a core organizing principle for the Workgroup=s collective work together. Further, EPA

expects to make development and use of the Exchange Network a core functional priority of its strategic investments, with the expectation that a growing number of States have and will continue to do so. We expect these joint investments to produce representative Exchange Network flows this year. We will use this experience to create a long-term implementation plan by the end of 2000.

## Appendix C: Acronym List

| | |
|---|---|
| **ADR** | **Active Data Retrieval** |
| **AIRS** | **Aerometric Information Retrieval System** |
| **CBI** | **Confidential Business Information** |
| **CDX** | **Central Data exchange** |
| **CIO** | **Chief Information Officer** |
| **CROMERR** | **Cross-Media Electronic Reporting and Records Rule** |
| **DET** | **Data Exchange Template** |
| **DMR** | **Discharge Monitoring Report** |
| **DSC** | **Data Standards Council** |
| **EDI** | **Electronic Data Interchange** |
| **EPA** | **United States Environmental Protection Agency** |
| **ECOS** | **Environmental Council of the States** |
| **FAQ** | **Frequently Asked Questions** |
| **FGDC** | **Federal Geographic Data Committee** |
| **FITS** | **Facility Identification Template for States** |
| **HTML** | **Hypertext Markup Language** |
| **HTTP** | **Hypertext Transfer Protocol** |
| **I-3** | **Information Integration Initiative** |
| **IDEF`** | **Interim Data Exchange Format** |
| **IETF** | **Internet Engineering Task Force** |
| **IMWG** | **State/EPA Information Management Workgroup** |
| **NA** | **Network Administrator** |
| **NEPPS** | **National Environmental Performance Partnership System** |
| **NPDES** | **National Pollution Discharge Elimination System** |
| **PCS** | **Permit Compliance System** |
| **PKI** | **Public Key Infrastructure** |
| **PPA** | **Performance Partnership Agreement** |
| **PPG** | **Performance Partnership Grant** |
| **RCRIS** | **RCRA Information System** |
| **SEA** | **State-EPA Agreement** |
| **SHTTP(s)** | **Secure Hypertext Transfer Protocol** |
| **SIC** | **Standard Industrial Classification** |
| **SOAP** | **Simple Object Access Protocol** |
| **SSL** | **Secure Socket Layer** |
| **STORET** | **Water quality information system** |
| **TCP/IP** | **Transmission Control Protocol/Internet Protocol** |
| **TPA** | **Trading Partner Agreement** |
| **TPAmL** | **Trading Partner Agreement Markup Language** |
| **VPN** | **Virtual Private Network** |
| **WWW** | **World Wide Web** |
| **XML** | **eXtensible Markup Language** |
| **XQL** | **Extensible Query Language** |

## Appendix D: Network Blueprint Glossary

### Network Components

**Data Exchange Templates –** empty but defined templates for data presentation and exchange. They identify what types of information are required for a particular document (i.e., name, address, etc.) as established in predefined standards or agreements.

**Data Standards –** "documented agreements on formats and definitions of common data.", according to the Environmental Data Standards Council.

**Node –** a participant's single, managed portal for providing and receiving information via the Network.

**Stewardship –** the management of Network assets in order to ensure their accessibility and integrity.

**Technical Infrastructure –**the software, hardware and protocols use to make the Network function.

**Trading Partner Agreement** – an agreement in the form of documents formally adopted by two or more partners for the purpose of defining the responsibilities of each party, the legal standing (if any) of the proposed exchange, and the technical details necessary to initiate and conduct electronic information exchange.

### Network Terminology

**Active Data Retrieval -** Within the Network and via CDX, EPA will use active data retrieval to obtain environmental data from other network nodes.

**CDX (Central Data eXchange)** – a centralized electronic report receiving system that will serve as EPA's enterprise-wide portal to the National Environmental Information Exchange Network. .

**CROMERRR (Cross Media Electronic Reporting and Records Rule)** – proposed rule issued by U.S. EPA in July 2000 that sets forth criteria for voluntary electronic environmental reporting and recordkeeping and intends to enable electronic submission of any document that the regulated community must submit or maintain under federal environmental laws.

**Data Element –** individual pieces of data that are standardized through common definitions and formats (data standards) (e.g., facility name).

**Data Stewardship** - Managing data, resources and activities including quality assurance, data collection, maintenance and disposition.

**Environmental Data Standards Council (EDSC**) – an independent forum established by the State/EPA Information Management Workgroup where States, Tribes whose mission is to promote the efficient sharing of environmental information between EPA, States, Tribes and other parties through the development of data standards.

**Network Portal –** point of entry into the Network that is established by each member through a common protocol. This may include links to web sites or search engines.

**Non-repudiation –** a service that provides proof of the integrity and origin of data, which can be verified by any third party at any time.

**Port –** a system that translates a piece of software to bring it from one type of computer system to another.

**Portal Authority** - the entity that controls access to a portal and maintains the web sites or search engines associated with that portal.

**State/EPA Information Management Workgroup (SEIMWG) –** a group jointly established by U.S. EPA and the Environmental Council of the States (ECOS) whose mission is to improve the collection, management, and use of environmental data through providing a forum for resolving information issues between states and EPA; learning from each other's efforts and investments; and achieving a shared vision of future environmental information management.

**Transaction** – document template containing data, including common header and footer information, exchange network standard data elements, and program/flow specific elements.

**Transmission –** one or more transactions moved across the exchange network

## Security Terms

**Authentication -** process of verifying the identity of the sender and the integrity of the message. This can be done through the use of SSL, PKI, or other mechanisms.

**Digital Certificate –** a record that is used to establish a secure connection. It contains information about who it belongs to, who it was issued by, a unique serial number or other unique identification, valid dates, and an encrypted "fingerprint" that can be used to verify the contents of the certificate.

**Electronic Signature –** an electronic record usually attached to a larger record that is used by an individual as the legal equivalent of a handwritten signature.

**PKI (Public Key Infrastructure)** – a system for issuing and validating digital certificates, including a root certificate authority a certificate repository or directory, a certificate practice statement and trained individuals performing trusted roles to operate and maintain the system.

**SSL (Secure Sockets Layer)** – a protocol designed by Netscape Communications to enable encrypted, authenticated communications across the Internet. Users on both sides are able to authenticate data and ensure message integrity.

## Technical Terms

**EDI** (electronic data interchange) – the transmission of information between computers

**FTP (File Transfer Protocol)** - tool used to transfer files through the Internet from one computer to another

**HTML (HyperText Mark up Language) –** coding language of data standards that indicates how to format text exchanged electronically. A block of text is surrounded with codes that indicate how it should appear.

**HTTP (Hypertext Transfer Protocol**) - A set of rules for moving hypertext files across the Internet

**Metadata –** "data about data" that describe the content, quality, condition, and other characteristics of data. Metadata accompanies the data set through its transmission.

**SHTTP(s) (Secure Hypertext Transfer Protocol**) **-** HTTP with the addition of security using Secure Sockets Layer.

**TCP/IP (Transmission Control Protocol/Internet Protocol)** – the suite of protocols that defines the Internet

**XML (eXtensible Markup Language**) – electronic language that expresses and transports data standards and transaction sets. XML uses an extensible set of tags to describe the meaning of data.

**VPN (Virtual Private Network)** – A network that can be run over the public Internet while still giving privacy and/or authentication to each user of the network.

## Other

**NEPPS (National Environmental Performance Partnership System)** – a joint state/EPA system established in 1995 that allows states and tribes greater flexibility and control in managing environmental programs.

**PPA (Performance Partnership Agreement) –** a broad strategic document containing a joint statement of priorities and goals negotiated between a state and EPA Region. Sometimes called an Environmental Performance Agreement.

**SEA (State/EPA Agreement) –** annual operating agreements negotiated between states and their EPA Regions.

## Appendix E: Complex Data Standard Example

### PROPOSED FACILITY IDENTIFICATION DATA STANDARD
### FINAL DRAFT - JULY 19, 1999 (Revised Draft 7/21/99)

This standard describes the data elements used to uniquely identify a facility site and differentiate it from other facility sites. The standard was developed by a state/EPA Action Team (team) chartered by the ECOS-State-EPA Information Management Workgroup. The standard provides guidance to those developing systems to manage facility data and to data trading partners who wish to exchange facility identification information.

"Facility" and "Site" are terms that have been defined differently in various environmental regulations and programs. This standard relies on the common English definitions of facility and site, in order to accommodate multiple technical definitions of both terms.

The data elements are organized into groups (i.e., Facility Site, Geographic Coordinates, Affiliation, Organization, Individual, Mailing Address, Environmental Interest, Standard Industrial Classification, and North American Industry Classification). Each group represents a different thing of significance, related to the identification of a facility site, about which information needs to be known. For each group, a definition is provided, and a specification of its relationships to the other groups. Facility Site is the central group, and all other groups are related to it either directly or indirectly.

This standard does not establish new or modify existing data collections, reporting requirements or system development requirements. This standard addresses data elements most relevant to the identification of a facility site. The team also agreed that the data elements which describe basic information about the facility site and the environmental interests associated with the facility site are always necessary. The standard allows trading partners and data managers to establish rules about the specific elements they need to manage, as their business needs dictate. It also avoids any confusion that might result from elements prescribed as "mandatory" being interpreted by others as establishing or modifying a data collection or reporting requirement.

The team recognizes that agencies (both state and EPA) will designate elements and relationships as mandatory under specific circumstances and that implementation of this standard in existing or future data exchanges will require the development and/or negotiation of any programmatic changes or exchange-specific rules. The standard is not intended to represent a minimum nor a maximum set of data that an agency should collect, manage or exchange to meet its facility identification business needs. These are implementation issues outside the scope of the standard itself.

A definition and format is given for each data element. The format provides the maximum length of the data element and the data type. Data types include alphanumeric (A), number (N), and Date. For several data elements, allowable values are also included. In most cases, the allowable values shown are provided as examples to illustrate the intended use of their respective element. These are noted as "examples" in the text. In a few cases, however, where the standard makes use of a reference set defined elsewhere (e.g. FIPS codes) or where the set of allowable values is small and stable, the values shown or referenced represent the initial permitted code set for their respective elements.

| DATA ELEMENT NAME | DATA ELEMENT DEFINITION | FORMAT |
|---|---|---|
| **Facility Site**<br>Definition: Basic identification information for a facility site, including the facility registry identifier, geographic address, and geopolitical descriptors.<br>Relationships:<br>❑ Each Facility Site may be involved with one or more Affiliation(s).<br>❑ Each Facility Site may be classified by one or more Standard Industrial Classification(s).<br>❑ Each Facility Site may be classified by one or more North American Industry Classification(s).<br>❑ Each Facility Site may be geographically located by one or more Geographic Coordinates.<br>❑ Each Facility Site must be monitored by one or more Environmental Interest(s). | | |
| Facility Registry Identifier | The identification number assigned by the EPA Facility Registry System to uniquely identify a facility site. | A(12) |
| State Facility Identifier | The unique identification number used by a state to identify a facility site. | A(12) |
| Facility Site Name | The public or commercial name of a facility site (i.e., the full name that commonly appears on invoices, signs, or other business documents, or as assigned by the state when the name is ambiguous). | A(80) |
| Location Address | The address that describes the physical (geographic) location of the front door or main entrance of a facility site, including urban-style street address or rural address. | A(50) |
| Supplemental Location Text | The text that provides additional information about a place, including a building name with its secondary unit and number, an industrial park name, an installation name or descriptive text where no formal address is available. | A(50) |
| Locality Name | The name of the city, town, village or other locality, when identifiable, within whose boundaries (the majority of) the facility site is located. This is not always the same as the city used for USPS mail delivery.<br><br>Allowable Values: (examples)<br>"None" is an allowable value.<br>The code set found in the current FIPS 55 Guideline: Codes for Named Populated Places, Primary County Divisions, and Other Locational Entities of the United States, Puerto Rico, and the Outlying Areas. The URL is:<br>http://www.itl.nist.gov/div897/pubs/fip55-3.htm. | A(60) |
| County and State FIPS Code | The code that represents the county or county equivalent and the state or state equivalent of the United States.<br>Allowable Values:<br>All codes for counties and county equivalents of all states of the U.S. as well as U.S. territories and possessions found in the current FIPS publication 6-4, Counties and Equivalent Entities of the United States, Its Possessions, and Associated Areas.<br>Remarks: The first 2-digits of the code represent the state; the last 3-digits represent the county. For example, 09001 represents Fairfield County (001), Connecticut (09). | A(5) |

| DATA ELEMENT NAME | DATA ELEMENT DEFINITION | FORMAT |
|---|---|---|
| County Name | The name of the U.S. county or county equivalent in which the facility site is physically located. | A(35) |
| State Name | The name of a principal administrative subdivision of the United States, Canada, or Mexico. | A(35) |
| Country Name | The name that represents a primary geopolitical unit of the world. Default: United States | A(44) |
| Location ZIP Code/ International Postal Code | The combination of the 5-digit Zone Improvement Plan (ZIP) code and the four-digit extension code (if available) that represents the geographic segment that is a subunit of the ZIP Code, assigned by the U.S. Postal Service to a geographic location; or the postal zone specific to the country, other than the U.S., where the facility site is located. | A(14) |
| Tribal Land Name | The name of an American Indian or Alaskan native area where the facility site is located. | A(52) |
| **Geographic Coordinates**<br>Definition: A geographic point, or set of points, defined by latitude and longitude coordinates used to locate a facility site, usually the front door or centroid, including the associated method, accuracy, and description data.<br>Relationships: Each Geographic Coordinates occurrence must geographically locate one and only one Facility Site occurrence.<br>Remarks: This group is included by reference to the EPA Latitude/Longitude Data Standard; only mandatory data elements are shown. For an example allowable values list, see the Environmental Data Registry for EPA's list (URL: http://www.epa.gov/edr/). There can be multiple Geographic Coordinates associated with a Facility Site, however, each instance of a Geographic Coordinate can only be associated with one Facility Site occurrence. | | |
| Latitude Measure | The measure of the angular distance on a meridian north or south of the equator. | A(6) - A(10)<br>DD.dddddd |
| Longitude Measure | The measure of the angular distance on a meridian east or west of the prime meridian. | A(7) - A(11)<br>DDD.dddddd |
| Horizontal Accuracy Measure | The measure of the accuracy (in meters) of the latitude and longitude coordinates. | A(6)<br>in meters |
| **Geometric Type** (Textual Data or Code Data acceptable) | | |
| Code | The code that represents the geometric entity represented by one point or a sequence of latitude and longitude points. | A(3) |
| Name | The name that identifies the geometric entity represented by one point or a sequence of latitude and longitude points. | A(6) |
| **Horizontal Collection Method** (Textual Data or Code Data acceptable) | | |
| Code | The code that represents the method used to determine the latitude and longitude coordinates for a point on the earth. | A(3) |
| Text | The text that describes the method used to determine the latitude and longitude coordinates for a point on the earth. | A(60) |
| **Horizontal Reference Datum** (Textual Data or Code Data acceptable) | | |

| DATA ELEMENT NAME | DATA ELEMENT DEFINITION | FORMAT |
|---|---|---|
| Code | The code that represents the reference datum used in determining latitude and longitude coordinates. | A(3) |
| Name | The name that describes the reference datum used in determining latitude and longitude coordinates. | A(7) |
| **Reference Point** (Textual Data or Code Data acceptable) | | |
| Code | The code that represents the place for which geographic coordinates were established. | A(3) |
| Text | The text that identifies the place for which geographic coordinates were established. | A(60) |
| Source Map Scale Number | The number that represents the proportional distance on the ground for one unit of measure on the map or photo.<br>Remarks: Mandatory for all horizontal data collection methods except for methods using Global Positioning System (GPS). | A(10) |
| **Affiliation**<br>Definition: The relationship between a facility site and an organization and/or an individual person.<br>Relationships:<br>❑ Each Affiliation occurrence must be established with one and only one Facility Site occurrence.<br>❑ Each Affiliation occurrence may involve one and only one Organization occurrence.<br>❑ Each Affiliation occurrence may involve one and only one Individual occurrence.<br>❑ Each Affiliation occurrence may receive mail at one and only one Mailing Address occurrence.<br>Remarks: This doesn't imply that the affiliation must exist, but; if an affiliation exists, it must be associated with one and only one Facility Site. An Organization may participate with multiple Facility Sites. Each type of affiliation can exist more than once for a facility site; thus there can be two occurrences of the Legally Responsible Entity affiliation type with a Facility Site. | | |
| Affiliation Type | The name that describes the capacity or function that an organization or individual serves for a facility site.<br>Allowable Values (examples):<br>Organization<br>Legally Responsible Entity<br>Legal Operator<br>Waste Treater<br>Waste Handler<br>Land Owner<br>Parent Corporation<br>Individual<br>Report Certifier<br>Regulatory Contact<br>Public Contact | A (40) |
| **Organization**<br>Definition: A company, government body, or other type of organization that has some responsibility or role at the Facility Site.<br>Relationships: Each Organization must be involved with one or more Affiliation(s). | | |
| Organization Formal Name | The legal, formal name of an organization that is affiliated with the facility site. | A(80) |
| Organization DUNS Number | The Data Universal Numbering System (DUNS) number assigned by Dun and Bradstreet to identify unique business establishments. | A(9) |
| **Individual**<br>Definition: An individual person who has some responsibility or role at the facility site.<br>Relationships: Each Individual must be involved with one or more Affiliation(s). | | |
| Individual Full Name | The complete name of a person, including first name, middle name or initial, and surname. | A(70) |

| DATA ELEMENT NAME | DATA ELEMENT DEFINITION | FORMAT |
|---|---|---|
| Individual Title Text | The title held by a person in an organization. | A(40) |
| **Mailing Address**<br>Definition: The standard address used to send mail to an individual or organization affiliated with the facility site.<br>Relationships: Each Mailing Address must be the delivery point for one or more Affiliation(s). | | |
| Mailing Address | The exact address where a mail piece is intended to be delivered, including urban-style street address, rural route, and PO Box. | A(50) |
| Supplemental Address Text | The text that provides additional information to facilitate the delivery of a mail piece, including building name, secondary units, and mail stop or local box numbers not serviced by the U.S. Postal Service. | A(50) |
| Mailing Address City Name | The name of the city, town, or village where the mail is delivered. | A(30) |
| Mailing Address State Name | The name of the state where mail is delivered. | A(35) |
| Mailing Address Country Name | The name of the country where the addressee is located.<br>Default: United States | A(44) |
| Mailing Address ZIP Code/International Postal Code | The combination of the 5-digit Zone Improvement Plan (ZIP) code and the four-digit extension code (if available) that represents the geographic segment that is a subunit of the ZIP Code, assigned by the U.S. Postal Service to a geographic location to facilitate mail delivery; or the postal zone specific to the country, other than the U.S., where the mail is delivered. | A(14) |
| **Environmental Interest**<br>Definition: The environmental permits and regulatory programs that apply to the facility site.<br>Relationships: Each Environmental Interest occurrence must apply to one and only one Facility Site occurrence. | | |
| Environmental Interest Type | The environmental permit or regulatory program that applies to the facility site.<br>Allowable Values: (examples)<br>Value — Meaning<br>TRI Report — Toxic Release Inventory Report<br>NPDES Major — Clean Water Act NPDES Major<br>NPDES Minor — Clean Water Act NPDES Minor<br>NPDES General Permit — Clean Water Act NPDES General Permit<br>NPDES Stormwater Permit — Clean Water Act NPDES Stormwater Permit<br>NPDES Other Permit — Clean Water Act NPDES Other Permit<br>Pretreatment SIU — Clean Water Act Pretreatment Significant Industrial User<br>SS Major — Clean Air Act Stationary Source Major<br>SS Synthetic Minor — Clean Air Act Stationary Source Synthetic Minor<br>SS NESHAP Minor — Clean Air Act Stationary Source NESHAP Minor<br>SS Other Minor — Clean Air Act Stationary Source Other Minor<br>RMP Facility — Clean Air Act RMP Facility<br>TSD — Hazardous Waste Handler - Treatment, Storage, Disposal (TSD)<br>LQG (EPA Defined) — Hazardous Waste Handler - Large Quantity Generator (LQG), as defined by EPA<br>LQG (State Defined) — Hazardous Waste Handler - Large | A(60) |

| DATA ELEMENT NAME | DATA ELEMENT DEFINITION | FORMAT |
|---|---|---|
| | Quantity Generator (LQG), as defined by State<br>SQG Hazardous Waste Handler - Small Quantity Generator (SQG)<br>Spill Control Plan Oil Pollution Act Spill Control Plan<br>UIC Underground Injection Control Well (UIC)<br>UST Underground Storage Tank (UST)<br><br>Remarks: This list will be expanded as necessary to include values for additional interests. | |
| Environmental Interest Start Date | Date the agency became interested in the facility site for a particular environmental interest type. | Date |
| Environmental Interest End Date | Date the agency ceased to be interested in the facility site for a particular environmental interest type. | Date |
| Environmental Interest Start Date Qualifier | The qualifier that specifies the meaning of the date being used as an approximation for the environmental interest start date.<br>Allowable Values: (examples)<br>Date of First Report<br>Date Operations Commenced<br>Date of Permit Application<br>Date Permit Issued<br>Date Monitoring Started | A(50) |
| Environmental Interest End Date Qualifier | The qualifier that specifies the meaning of the date being used as an approximation for the environmental interest end date.<br>Allowable Values: (examples)<br>Date of last report<br>Date Permit Ended<br>Date Operations Ended | A(50) |
| Environmental Information System Abbreviated Name | The abbreviated name that represents the name of an information management system for an environmental program. | A(15) |
| Environmental Information System Identification Number | The identification number, such as the permit number, assigned by an information management system that represents a facility site, waste site, operable unit, or other feature tracked by that Environmental Information System. | A(30) |
| **Standard Industrial Classification**<br>Definition: The Standard Industrial Classification (SIC), or type of business activity, occurring at the facility site.<br>Relationships: Each Standard Industrial Classification occurrence must classify one and only one Facility Site occurrence.<br>Remarks: This group is included by reference to the SIC/NAICS Data Standard. | | |
| Standard Industrial Classification (SIC) Code | The code that represents the economic activity of a company (4-digits). | A(4) |
| SIC Primary Indicator | The name that indicates whether the associated SIC Code represents the primary activity occurring at the facility site.<br>Allowable Values:<br>Value Meaning<br>Primary The SIC Code represents the primary activity occurring at the facility site.<br>Secondary The SIC Code represents a secondary activity occurring at the facility site.<br>Unknown It is not known whether the SIC Code represents | A(10) |

| DATA ELEMENT NAME | DATA ELEMENT DEFINITION | FORMAT |
|---|---|---|
| | the primary or secondary activity at the facility site. | |
| **North American Industry Classification**<br>Definition: The North American Industry Classification System (NAICS) code, or type of industrial activity, occurring at the facility site.<br>Relationships: Each North American Industry Classification must classify one and only one Facility Site.<br>Remarks: This group is included by reference to the SIC/NAICS Data Standard. | | |
| North American U.S. National Industry Classification System (NAICS) Code | The code that represents a subdivision of an industry that accommodates user needs in the United States (6-digits). | A(6) |
| NAICS Primary Indicator | The name that indicates whether the associated NAICS Code represents the primary activity occurring at the facility site.<br>Allowable Values:<br>Value Meaning<br>Primary The NAICS Code represents the primary activity occurring at the facility site.<br>Secondary The NAICS Code represents a secondary activity occurring at the facility site.<br>Unknown It is not known whether the NAICS Code represents the primary or secondary activity at the facility site. | A(10) |

**Appendix F:**

# TRADING PARTNER AGREEMENT

**BETWEEN THE [PARTNER NAME]**
**AND THE [PARTNER NAME]**
**FOR THE PARTICIPATION IN THE**
**[NAME OF DATA EXCHANGE PROGRAM]**

## I. PURPOSE

The purpose of this Trading Partner Agreement (TPA) is to identify the activities that 'PARTNER NAME' will undertake as a partner of the [NAME OF DATA EXCHANGE PROGRAM]. As a [NAME OF DATA EXCHANGE PROGRAM] Partner, [PARTNER NAME] will work cooperatively with [OTHER PARTNER] NAME' to design and implement a [DESCRIBE DATA EXCHANGE] which will make data and information readily available to [WHOM DATA IS AVAILABLE TO], using Internet technology.

## II. BACKGROUND

The [PARTNERS] are – [Federal, State, regional and local governments, non-profit, private industry, academic, and private citizens] -- dedicated to environmental protection. … These organizations have interrelated missions for the production and exchange of data and information to guide management decisions and practices that affect [DESCRIBE WHY THIS DATA EXCHANGE IS NECESSARY OR DESIRED].

## III. BENEFITS OF EXCHANGE

The most direct benefit of this shared information resource will be access to more timely information, in a commonly accessible format. By publishing data and information on the Internet, the Partners can also expect:

[LIST THE BENEFITS TO BE GAINED BY ALL PARTNERS, EXAMPLES INCLUDE:]

- Substantial improvements in locating and using information among offices within the Partners' own agency. This will make internal information more readily available between offices within an agency.
- Data producers will have ownership of the data they generate. They can control the quality and timeliness of the data. There will no longer be several 'copies' of data that each has different levels of quality.

- Savings from reduced staff labor responding to information requests. A 75% reduction in cost responding to user requests has been realized by the [PARTNER OR OTHER ORGANIZATION] to date.
- Reduced dependence on paper files that get out-of-date.
- Better quality information by using shared policies and guidelines. Standardized policies and guidelines are imperative for a distributed system to function smoothly.
- Designs created by the [PARTNERS] will substantially reduce the cost of developing similar systems by other organizations.
- Database and software tools created by the [PARTNERS] will substantially reduce the cost of developing similar tools by other organizations.
- Standardized designs and tools will substantially reduce the cost of integrating, analyzing, modeling, and reporting information.
- Better marketing of agency capabilities inside and outside the agency. This is very important to public agencies when it is time to justify the budget.

## IV. PARTNERS' ROLES AND RESPONSIBILITIES

Participation in [PROGRAM] requires good-faith effort by each partner to make a distributed information system function as efficiently as possible. Due to the loosely structured nature of a distributed information network, each partner must take absolute responsibility for working cooperatively with other Partners to make data and information more readily available for online Internet access.

To ensure compliance with this agreement, the partners must demonstrate in their Workplan that they have the management commitment, adequate hardware, software, and network infrastructure, and technical resources to conduct the level of participation in [PROGRAM] that they desire. Participation does not supersede any data or information management and reporting requirements of any grant or contract.

As a successful Partner, data or information published over the Internet may serve as an official deliverable if all grant or contract requirements are met and the funding agency project officer or contract officer has agreed to this practice in writing. Formal notification of any published deliverables must be made directly to the project officer or contract officer so that there is an official record of the deliverable by the due date. In selected cases, the [PARTNERS] will maintain a 'mirror' copy of data and information published on a separate server by a Partner. The purpose of the mirror copy is to make a duplicate copy available in case of network failure and to make off-site copies in case of catastrophic failure. Partners are encouraged to design their information management systems so that their information, if appropriate, can be mirrored to the [PROGRAM] servers on a regular basis.

- PUBLICATION
- ACCESS to information, servers, software, and staff resources developed policies and guidelines …
- DATA STEWARDSHIP

- SECURITY
- DESIGN STANDARDS
- TECHNICAL GUIDANCE DOCUMENTS
- PUBLIC AND TECHNICAL OUTREACH OR TRAINING
- COORDINATION OF TECHNOLOGY OR DATA ISSUES
- COSTS AND RESOURCES

## V. AUTHORITIES AND POLICIES

The [PROGRAM] is founded on the commitment of [PARTNERS] to work in partnership to protect and restore the [DESCRIBE]. This TPA specifies the roles and responsibilities for maintaining key information for public access and derives its authority from, and is a supplement to the [NOTE POLICY OR AUTHORITY].

1. This TPA will comply with existing [DESCRIBE] policies with regard to [DESCRIBE MANDATES] for all data exchanged under the auspices of this agreement.

2. ANY COSTS ASSOCIATED WITH PUBLICATION OR DISTRIBUTION OF DATA

3. NECESSARY COMPLIANCE WITH DATA STANDARDS INCLUDING METADATA

4. GUIDELINES IDENTIFY THE METADATA FIELDS THAT ARE ESSENTIAL FOR SEARCHING, LOCATING, QUERYING, AND RETRIEVING DATA AND INFORMATION BY THE [PROGRAM] INTERFACE, WHICH WILL GIVE USERS EASIER ACCESS TO INFORMATION FROM VARIOUS PARTNERS.

**TRADING PARTNER AGREEMENT**

The 'PARTNER NAME' agrees to the following:

A. INFORMATION ACCESS (Choose one of the following)

❒ Shall operate and maintain a publicly accessible Internet web site ('WEB SITE NAME') which is continuously operational.

❒ Shall maintain data and information on a publicly accessible Internet web site('WEB SITE NAME') which is continuously operational.

❒ Shall provide data and information for publication on a publicly accessible Internet website('WEB SITE NAME') which is continuously operational.

❒ Other _____.

B. METADATA (Choose one of the following)

❒ Shall provide metadata (meeting specifications documented in [PROGRAM] *Metadata Reporting Guideline*s) and maintain a metadata base linked to each data set published on the Internet.

❒ Shall provide metadata (meeting specifications documented in [PROGRAM] *Metadata Reporting Guideline*s) to be linked to each data set we have published on the Internet.

❒ Other _____.

C. STANDARDS (Choose those that are applicable)

❒ Shall comply with all [PROGRAM] data management policies and guidelines and will participate in the [DATA EXCHANGE DESCRIPTION] standards development / modification process.

❒ Shall comply to the extent possible with [PROGRAM] data management policies and guidelines. Exceptions shall be documented in the Workplan, Section X, and in metadata files, as appropriate.

❒ Shall meet the following specific practices, to the extent possible:
- ❒ Submit deliverables in electronic format.
- ❒ Meet or exceed Locational Accuracy, ___ meters.
- ❒ Report locations in Longitude/Latitude decimal degrees, ___ decimal accuracy.

- ❒ Report locations in Coordinate Projection UTM Zone 18, ___ meters accuracy.
- ❒ Report locations in North American Datum (NAD83).
- ❒ Report locations in North American Vertical Datum (NAVD88).
- ❒ Adhere to [PROGRAM] Metadata Reporting Guidelines.
- ❒ Adopt the [PROGRAM] Naming Conventions.
- ❒ Adopt the [PROGRAM] Data Dictionary.
- ❒ Adopt [PROGRAM] database designs to encourage standardization between agencies' databases by using shared designs.
- ❒ Adopt the [PROGRAM] Calendar Date Policy.
- ❒ Adopt the [PROGRAM] Method Codes.
- ❒ Adopt the [PROGRAM] Numeric Data Reporting Guideline.
- ❒ Other _____.
- ❒ Exceptions _____

D. DATA and INFORMATION (Choose those that are applicable)

❒ Shall make all pertinent, non-sensitive data and information publicly available through Internet access.

❒ Shall make all federally and state funded or match grant and contract, non-sensitive data and information publicly available through Internet access.

❒ Shall make all federally funded or match grant and contract, non-sensitive data and information publicly available through Internet access.

❒ Shall make selected, non-federally funded or match grant and contract, data and information publicly available through Internet access.

❒ Intend to 'mirror' web site contents ('WEB SITE NAME') to the [PROGRAM] servers.

❒ Intend to 'mirror' data and metadata to the [PROGRAM] servers.

❒ Other _____.

E. TECHNOLOGY

❒ Shall utilize [TYPE] technology in exchange

[LIST APPROPRIATE HARDWARE, SOFTWARE, PLATFORMS, ETC.]

❒ Other _____.

F. SECURITY

❒ Shall utilize [TYPE] technology security

❒ Shall follow [TYPE] security policies and procedures

[LIST APPROPRIATE TECHNOLOGIES AND PROCECURES]

❒ Other _____

## VI. FINANCIAL ARRANGEMENTS

It shall be the responsibility of each Partner to secure the resources required to meet the requirements of this TPA. It shall be the responsibility of parties negotiating an exchange of data or information to address any financial requirements associated with any such exchange. The Partners will not charge a fee for data or information published over the Internet.

## VII. PERIOD OF AGREEMENT and TERMINATION

This TPA becomes effective on the date of signatures by both parties and continues until modified by mutual consent or unless terminated with 60 days written notice by any party. This TPA should be reviewed annually and amended or revised when required.

## VIII. DATA OWNERSHIP AND RIGHTS

Any data or information and accompanying metadata that are [PROGRAM] grant or contract deliverables managed under this TPA, shall be transferred to the [PARTNERS] in the event that the organization can no longer meet the requirements of this MOA.

## IX. POINTS OF CONTACT

The following individuals have been identified as points of contacts for the roles and responsibilities defined:

| Role/Responsibility | Point of Contact | Phone | E-Mail |
|---|---|---|---|
| Publication<br>Access to Information<br>Data Stewardship<br>Design Standards<br>Technical Guidance Documents<br>Public and Technical Outreach<br>Coordination of Technology or Data Issues<br>Costs and Resources | | | |

## X. WORKPLAN

Provide a descriptive plan of the resources that are available (management, hardware, software, networks, staff), how those resources will be utilized, the scope of activities that you will conduct, your agency's plan for quality assurance and quality control, and any other pertinent information. Any plans for electronic publication of grant or contract deliverables must include a very clear description of the process, reporting, and sign off. This workplan provides your agency the opportunity to personalize its involvement in CIMS.

## XI. APPROVALS

_____

COMMISSIONERS, AGENCY DATE

_____

COMMISSIONER, AGENCY DATE

**THE EXCHANGE NETWORK**
**A White Paper of the**
**INFORMATION INTEGRATION INITIATIVE**
Draft - August 1, 2000[1]

## PURPOSE

This paper intends to advance the discussion on the opportunities for state environmental agencies and EPA in implementing an Exchange Network by: 1) discussing the need for and benefits of the Exchange Network concept; 2) defining the components of the Exchange Network, and 3) discussing strategic implications and implementation issues, and 4) making recommendations for action.

## DISCUSSION

**Background:** In 1998, the State/EPA Information Management Workgroup (S/E IMWG) proposed a vision and core operating principles for creating a partnership for collaborative environmental information management[2]. Since then, a more specific vision for how this partnership might be realized in the form of a national environmental information exchange network (Exchange Network) has been evolving. The Exchange Network vision is one where participating agencies avail their information holding to other participants of the Exchange Network directly from their own agency's web presence[3], based on agreed-upon neutral standards-based formats and secure Internet transaction protocols. (Detailed discussion of the shared expectations for the Exchange Network and potential implementation steps can be found in Attachments A and B).

**Why the Exchange Network?:** Three primary drivers are evolving that make it essential that environmental regulatory agencies re-think the information management infrastructure they employ to collect, use, and share environmental data:

1**) The changing nature of state and federal environmental protection roles:** A wide array of individual information-sharing relationships exists between states and EPA. Each individual information-sharing relationships was designed to meet specific business needs and state and federal legislative demands. As the demand for integrated environmental information has risen, the collective complexity of these information sharing relationships has created a situation where information is difficult and burdensome to share across programs or organizational boundaries.

**2) The changing nature of the environmental protection business:** a) The business elements of environmental protection continue to face a growing emphasis on cross-media, integrated, results-based approaches to environmental protection, b) pressures from the regulated community to rationalize the environmental reporting process and reduce burden, and c) a legal and policy commitment to effective public access[4]. Thirdly,

**3) The increasing expectations of the American public** for government to follow the private sector's lead in implementing information technology to improve customer service and allow for transparent access to environmental information, regardless of which level of government is responsible for it. The success of private companies in using Internet-based technologies to cut costs and increase productivity has been attributed by some to the ability of company management to consider new business arrangements- new supply line models, and unconventional organizational relationships. The Agency should be equally creative and open to the possibility of change.[5]

**Benefits of the Exchange Network**:

Implementation of the Exchange Network effectively will create a "standards-based" lexicon of environmental information. This will have significant impacts to our many efforts to improve environmental protection. Retrofitting such an infrastructure in place will: 1) **improve** the capacity to conduct cross-media, integrated, results-based approaches to environmental protection; 2) **rationalize** the environmental reporting process and thus reduce reporting burden on the regulated community; and, 3) **allow** for improved understanding of the environmental information provided to regulators and the public by improving data quality, timeliness, and allowing for effective interagency error-correction processes. **Improvements in our ability to target resources to priority problems, to provide a more-informed policy-making process, to conduct cross-media impact assessment, and improve enforcement and compliance programs are all potential benefits of the Exchange Network.**

Adopting a neutral exchange format has many operational benefits: 1) it greatly simplifies and reduces the burden inherent in the current exchange processes; 2) it gets states out of the business of directly loading EPA national systems, solving state access problems and simplifying EPA information security control management; 3)dual data management and funding/resource concerns-dual data entry, dual quality concerns, dual error correction processes, etc. are minimized: and,
4) formats for data exchange can be based upon common business needs, rather than computer system design, and consequently can be consistent in format and style across media lines, allowing for a holistic change management system to be implemented. Significantly, once the exchange negotiation process is disconnected from system design, partners agencies are freed up to reengineer systems at their own pace without having to coordinate systems changes with regulatory partners. Consequently, states will be able to coordinate horizontally with other agencies within the state in response to the state CIO's directives on data and technology standards.

While ensuring EPA continued access to regulatory-required information collected in delegated state programs, and improving the efficiency of interagency information sharing, the Exchange Network also will offer many new opportunities. Network participants will be able to access and use many data collections not routinely exchanged between agencies. (i.e. PCS minors, UST, spatial data sets). States will be able to access each other's information collections as well as EPA's. Many new opportunities for collaborative public access strategies, that have not existed to date, can be explored and help us answer the question - How do we compliment each others public access offerings - and not duplicate them? More generally, the Exchange Network will allow for the collective exploration of opportunities to leverage each others assets, talents, and strengths.

**Components of the Exchange Network**

**Data Standards and Transaction Sets** - For common business areas where information is exchanged, a system of neutral exchange formats, composed of agreements on data content (data standards), data format (transaction sets), meta data, technical formats, quality specifications, and exchange schedules will be negotiated among participating agencies.

**Exchange Process** - Donor agencies will extract information from their internal systems, and host it on their Internet sites in the agreed-upon exchange format, where it will be available anytime for other partners to access. Where EPA is the receiving partner, EPA would acquire the information on a periodic schedule from the Internet, 'pull' it into the Central Data Exchange Facility (CDX), reformat the data from the exchange format into the program system-specified format, and load it into the EPA system[6]. Likewise, EPA would avail its information collections to states - the change in EPA's focus from Central Receiving to Central Data Exchange.

**Policy Infrastructure** - In order for the Exchange Network to operate, and be sustainable, an

interagency framework must be established to negotiate operational policies and business. Guidelines on data quality, timeliness, error correction, meta data expectations, and standard operating procedures will all need to be developed. Largely, the policy infrastructure can be guided by stewardship. Three delineations of stewardship can be assumed: 1) Network Governance - The Network itself will require interagency governance, people empowered to lead, manage, to establish and govern a framework for exchanges and trading partner agreements, and direct the expansion of the network beyond its initial participants; 2) Stewardship of the data itself will be required–the data standards, the transaction set standards, definitions, meta data etc. Each participating agency, as an Exchange Network partner, in agreeing to host their information assumes data management responsibility for their portion of the Exchange Net; and 3) Stewardship of the data exchange process. An active management system monitoring the operations and maintenance of the actual exchange will need to be established (Refer to the companion document--*Stewardship and Governance: A white paper of the Information Integration Initiative* for more detailed information.)

**Trading Partner Agreements** - A generic framework for how participating agencies share their information collections with others is required. In cases of regulatory reporting requirements, more specific and formal Trading Partner Agreements (TPA) will need to be negotiated. Currently information requirements are defined in many places (delegation agreements, NEPPS, ICRs, etc,) these will have to be coordinated.

**Technical Infrastructure** - Each participating agency will have to ensure that it can provide the capacity to offer access to its information holdings, while maintaining the security and integrity of their information systems. The private business-to business e-commerce sector is heavily investing in Extensible Markup Language (XML) for exchanging information between partners. To adopt XML as the preferred exchange protocol, technical issues (network capacity, security, Internet connections, changing versions of Internet protocols, browser upgrades) will all have to be examined for impact on participating agencies, as well as sustainability and stability of the network.

**Strategic Implications of the Exchange Network concept and Implementation Issues**

**Represents a new paradigm for sharing information** - The Exchange Network vision, where participants avail their information holding to other participants directly from their own agency's web presence represents a radical departure from the current state/EPA data 'reporting' relationship. Traditionally states have been responsible for directly loading information into individual EPA National systems. Using the Exchange Network, states would make their information available for EPA to access, and EPA would assume the responsibility for getting the information into its computer systems. Existing delegation agreements that specify information requirements, some NEPPS agreements, electronic reporting trading partner agreements, informal ad-hoc data acquisition arrangements, all will need to converge into documented Exchange Network trading partner agreements. This would also impact the information collection processes from the regulated community.

**Recognizes Interdependence** - While there has been a shift for most states from acting as agents for EPA to directly carrying the weight for the majority of environmental protection programs[7], our collective business functions remain inherently interdependent. As such, the information managers must collectively understand that since our business functions are interdependent, as are supporting information needs.

**Requires a Community** - The Exchange Network can only be successful if there is an interconnected community of people who exchange information via the Network. The network is not just the technical infrastructure and policies. It requires a functioning community of environmental regulators[8]. The network concept is going to require that this community work in ways it has traditionally not been accustomed to, and that will require leadership to achieve and support these new arrangements.

**Standards** - We will need to develop both data standards as well as standard transaction sets. Ideally the data standards would come first, but that may not be practical in many situations. Neither should necessarily hold the other up. While the Data Standards Council is operational, it is not positioned to take on the simultaneous negotiation of all the necessary transaction set formats. (The S/E IMWG should address this need.)

**Not everyone will be ready at the same time**. The concept of the Exchange Network requires that participants be current in information technology. For the foreseeable future, EPA will have to accommodate conducting business along traditional means for others still (or not) transitioning to the Exchange Network. This will require dual operation in many cases which will have resource implications.

**Trading Partner Agreements (TPA)** - Based upon our past experience and research into trading partner agreements[9], and following the e-commerce world, a mechanism for managing the many trading partner agreements will be necessary. It is essential that this be handled in a global manner to avoid many distinct individual trading partner agreements from being the norm. This process has to flush out issues such as unacceptable data quality, untimeliness, non-participation, and specify error-correction processes. Further, the TPA's would need to spell out for states managing federal programs, any other requirements unique to managing federal records (i.e. criminal enforceability)

**Technical Infrastructure** -Secure transactions- how to we ensure integrity of the network? Is a virtual private network (VPN) desirable? Can partners realistically post data outside firewalls for others to pick up or is through-the-firewall access going to be necessary? Levels of Internet traffic, readiness of XML, bandwidth requirements, and security measures all need careful investigation to ensure we can base our business arrangements upon them. EPA must ensure that its Central Data Exchange Facility (CDX) is capable of both receiving information from and providing information to the Exchange Network. We must better understand the data flows between agencies before committing to an implementation path. We must further research the technical path to EPA being able to "Come and get it" from states and mutually commit to a rational path to get there. One logical first step is to ensure that the on-going facility data synchronization pilots are successful and lead to implemented business practices.

**Policy Infrastructure** - What interagency business rules on the operations of the network should be established? Guidelines on data quality, timeliness, error correction, meta data expectations, etc. need to be negotiated? Change management practices can be synchronized. How do we best leverage each others work? Once initially set up, a second tier of issues will surface: What will the relationship to other external networks be (i.e. EDEN, Global Climate Change Net)? How should the network be broadened beyond initial participants? There may be pressure on EPA to broaden the Exchange Network faster than may be responsible. And thirdly, how can the negotiated exchange formats between agencies be leveraged to improve reporting streams from the regulated community?

## RECOMMENDATIONS

Interagency Recommendations**:**

- Establish a mechanism and process for negotiating exchange transaction formats
- A robust documentation of the current data flows and existing information trading agreements between states and EPA be carried out.
- Several pilots on information exchange between EPA and states should be started to isolate both technical and "business policy/practice problems, define solutions and implement "fixes" to start data exchanges. This begins with the successful completion of those currently underway (IDEF, Facility Exchange Pilots)
- A long-term implementation plan be developed by the end of FY2000.

EPA Recommendations

- To insure rapid agreement on the necessary standards the AA's must assign task force members from their programs to participate in interagency workgroups (i.e. the Data Standards Council) as a **priority Agency action**
- New ways to present meaningful integrated information to the public should be developed by examining and building on applications that show promise e.g., Chesapeake Bay Profiles, EnviroViz, Diana, Demographic Mapper (EJ), Decision Consequences Model--Region 3 and RAINS in Region 10.
- We should develop a truly integrated management and analysis system that integrates not only cross-program pollutant data but links it to pollution trends, GRPA results, ambient and facility compliance, enforcement actions, and our budget expenditures[10]
- We should be proactive in using our data to show how progress is being made under the each of the GPRA objectives to build cases, using data, to project future conditions and strengthen our budget requests with expected results and time frames.

**End Notes:**

1. August 1, 2000 draft reflects editing refinements only and no substantive changes from the July 27, 2000 draft.

2. *State/EPA Vision and Operating Principles for Environmental Information Management:* State/EPA Information Management Workgroup, January 1998. (www.state-epa-info-group.org/Vision/vision.html)

3. While the Exchange Network will involve many types of exchanges, the primary focus will be Internet-based and hence the this document is focused on Internet-based exchanges.

4. U.S. EPA, *The Problem with Environmental Reporting,* One Stop Reporting Program Strategy, 1996.

5. From *e-Government - An experiment in Interactive Legislation,* 2000. (http://cct.georgetown.edu/development/eGov/description.cfm)

6. All currently planned functions for CDX

7. US EPA & ECOS, *Environmental Pollutant Reporting Data in EPA's National Systems: Data Collection by State Agencies***,** June 1999.

8. "Letting go ... For a generation, highly centralized 'command and control' systems have been the primary means of managing the complex affairs of a community [enviro regulators] Now, following the private sector's lead, government is beginning to see that a more distributed approach, akin to that of a network, may be the better way to address the many messy, complex, and potentially competitive interrelationships that exist in a truly intelligent community" Peter Katz**,** *When Space & Time Collapse: The New Community*, Gov. Technology, May 2000

9. Extensive work on developing prototype Trading Partner Agreements has been done via the State Electronic Reporting/EDI Subcommittee (SEES) in conjunction with the National Governors' Association)

10. EnviroViz, Region-3 and RAINS, Region 10, have started down this path.

# Executable Trading-Partner Agreements in Electronic Commerce[1]

Martin Sachs[2], Asit Dan, Thao Nguyen, Robert Kearney, Hidayatullah Shaikh, Daniel Dias

IBM T. J. Watson Research Center
Yorktown Hts, NY 10598

## Abstract

In business to business electronic commerce, the terms and conditions describing the electronic interaction between businesses can be expressed as an electronic contract or trading-partner agreement (TPA) from which configuration information and code which embodies the terms and conditions can be generated automatically. This paper first discusses issues related to contracts and more generally to inter-business electronic interactions. Next, we describe the basic principles of electronic TPAs. The TPA expresses the rules of interaction between the parties to the TPA while maintaining complete independence of the internal processes at each party from the other parties. It represents a long-running conversation that comprises a single unit of business. Next, we describe our TPA language. We then describe tools for authoring TPAs and generating code from them. Finally, we describe an example of an application which can benefit from TPAs.

## 1. Introduction

Contracts describe legally enforceable terms and conditions in all kinds of interactions between people and organizations. Examples of interactions are marriage, employment, real estate purchases, and industrial supply arrangements. In business to business electronic commerce, there is a need to agree not only on the traditional terms and conditions but also on IT procedures from communication protocols to business protocols (Dan & Parr 1997a). Today, such contracts, or trading-partner agreements (TPAs), are generally written in human languages and then turned into code by programmers.

Business to business electronic commerce will be given considerable impetus by expressing the IT terms and conditions as electronic TPAs from which the code to perform the terms and conditions can be automatically generated at each party's business to business server. This will speed up the reduction of the terms and conditions to code and ensure that the code at each business partner's site will accurately embody the desired terms and conditions. In the longer term, electronic TPAs will also facilitate electronic negotiation of terms and conditions, at least for the simpler situations which need not involve extensive legal negotiation. Electronic negotiation in turn opens the possibility for spontaneous electronic commerce, i.e. quick and easy setup of business to business deals on the Internet (Dan *et al* 1998).

---



[2] Contact: mwsachs@us.ibm.com

In recent years, there has been a large amount of activity in modeling and analyzing various electronic commerce methods using contract or agreement approaches. Dan & Parr 1997b and Weigand & Ngu 1998 discuss how interoperable transactions in electronic commerce differ from traditional ACID (atomicity, consistency, isolation, durability) transactions (Gray & Reuter 1993) and the importance of distinguishing between the contract (communication behavior) and the task (the meaningful unit of work) and propose a scheme for specifying the contract which is suitable for analyzing the process.

Many academic publications discuss conceptual contracts as part of their models but the do not suggest a specific business to business contract language or discuss embodiment of a system based on such a contract. Dan & Parr 1997a discuss the general principles in business to business electronic commerce and mention the use of a business to business electronic contract but provide no details. Dan *et al* 1998 discuss the specific functions needed in a business to business electronic contract and describe the architecture of the prototype of a business to business server built at IBM Research but do not describe a specific contract language.
In this paper, we focus on the language for an electronic TPA and the tools to assist in composing the TPA and to generate code from it.

The paper is organized as follows. In section 2, we detail the issues that need to be addressed in business to business interactions. Section 3 discusses the principles of business to business electronic TPAs. In section 4, we describe our TPA language. In section 5, we describe the tools for creating TPAs and generating code from them. Finally, in section 6, we describe an application example which illustrates the use of the TPA.

## 2. Issues in Inter-Business Electronic Interactions

Increased automation of business processes within a business organization leads naturally to automation of business to business (B2B) interactions (Dan & Parr 1999). The issues of privacy, autonomy, heterogeneity in software and platforms, and more importantly, managing complexity of interactions, however, make this a challenging task. Some of these issues, e.g., heterogeneity of programming languages and platforms in which the application components are developed, and stateful interactions across program components, are also addressed in the automation of business internal processes and integrating application components. Total knowledge and control in the design of the business process within an organization make this a manageable task.

Component architectures such as CORBA (Corba 1998) and Enterprise Java Beans (Ejb 1999) provide middleware for integrating application components written in different languages. For the purpose of interaction, an application component needs to know only the interfaces to other components written in a suitable middleware integration language (e.g., Interface Definition Language or IDL in CORBA). In such environments, typically, the applications are executed as short ACID transactions. The underlying middleware provides necessary runtime services, e.g., naming, transaction, resource allocation. A long-duration application is modeled as a sequence of short independent steps invoked either manually or in an automated manner (Dan & Parr 1999,Wfmc 1998, Garcia-Molina & Salem 1987).

Most methodologies reported in the academic literature for automation of internal processes of individual businesses are not directly applicable for the automation of B2B interactions. First and foremost, no common shared underlying middleware can be assumed for distributed applications spanning organizational boundaries. Setting up such a common software bus requires tight coupling of the business partners' software platforms (e.g., consider the issues on security, naming, component registration).

Even if such a software bus can be established, ACID and/or complex extended transaction models of stateful interactions are not appropriate for such B2B interactions. First, implementation of such protocols necessitates tight coupling of operational states across business applications, which is highly undesirable. The application components in one organization may hold locks and resources in other organizations for an extended period of time, resulting in loss of autonomy. Rollback and/or compensation of application steps is no longer under the control of a single organization. Finally, in real-world business operations the states always move forward, and explicit recourse actions are taken by business partners to move to a more desirable operational state. An example is cancellation of a prior purchase or reservation.

In Dan and Parr 1997b, a conversational model of interactions is proposed where, based on the conversation history, each partner explicitly specifies the permissible operations. For management purposes, the internal business process is separated from external interactions. Each trading partner manages and is responsible for its own internal activities in the B2B application and may use ACID transactions within its own domain. The model furthermore structures the external interactions as actions consisting of requests, responses, modifications or cancellations, groups of actions that together satisfy certain interaction rules, and conversations demarcating interaction contexts. Interactions in one conversation may trigger actions in other conversations via execution of internal business logic.

The invocation of application components across organizational boundaries needs to be controlled and monitored (Dan and Parr 1997a, Dan *et al* 1998). First, without rigorous testing and cooperation in software development across organizations, the correct execution of such complex distributed applications can not be assumed. Second, in such automated interactions, trust becomes an overarching concern. During runtime, explicit checks are necessary to ensure that business partners are not violating any policy constraints (e.g., cancellation of a reservation must be within the allowable time window) .

In the Coyote (Cover Yourself Transaction Environment) project (Dan *et al* 1998), we address all of the above issues by setting up a B2B interaction via a composable interaction stack based on an electronic TPA. The automated process of setting up this interaction from an unambiguous formal specification and enforcing contractual agreements is termed an ***executable TPA***. The Coyote server provides additional services for supporting long running applications, e.g., application development, asynchronous event driven execution, compensation framework, maintaining correlation of conversations, logging and querying the activity on a conversation. However, these are not the focus of the current paper.

## 3. Principles of Business to Business Electronic TPAs

The purpose of the electronic TPA is to express the IT terms and conditions to which the parties to the TPA must agree in a form in which configuration information and the interaction rules which must be executable can be automatically generated from the TPA in each party's system. It should be understood that the information in the TPA is not a complete description of the application but only a description of the interactions between the parties. The application must be designed and programmed in the usual manner. As a simple example, the TPA may define requests such as "reserve hotel". The "reserve hotel" function must be designed, coded, and installed on the hotel server. That function may, in turn, invoke various site-specific functions and back-end processes whose details are completely invisible to the other party to the TPA.

We emphasize that the TPA is formulated to ensure that each party maintains complete independence from the other party both as to the details of the implementations and as to the nature of the business processes and back-end functions (database, transaction monitors, ERP functions, etc.) used. For example, as previously mentioned, the TPA neither requires, nor provides the means for, ACID transactions involving both parties.

In this paper, we use the terms "client" and "server" in the usual way. A client requests services of a server. However we envision applications in which a given party may play both server and client roles at different times. In other words, a party may both request services of the other party and receive service requests from the other party. In the simplest applications, there may be two parties, one of which is a always a server and the other, always a client. An example is a travel application involving a travel agency (client) and airline company (server). Even in such a simple case, however, the parties may exchange roles. For example, the airline company may issue requests to the travel agency for more information about the traveler or itinerary.

The TPA is represented at each party which acts as a server by an object, called a TPA object or (or equivalent code for non-object-oriented implementations), which performs rule checking and translation of the request messages from the form defined in the TPA to the actual method calls at the parties which act as servers. A similar TPA object, generated at each party which can act as a client to other party, performs the inverse translation, from local method calls to the request messages, as defined in the TPA, which are sent to the other party. A party which can act as both a client and as a server has both kinds of TPA object. Use of the TPA objects is illustrated in the examples in section 6.

The TPA represents a single long-running conversation, which is a set of related interactions, dispersed in time, that comprises a single unit of business. For example, in a travel application, the TPA might define the interactions between the travel agent and a hotel company starting with making the different reservations needed by the traveler, to the check-in processes during the trip, and ending when the traveler checks out at the last stop. This sequence of steps is a single long-running conversation. A unit of business is performed under the TPA by instantiating the TPA as a long-running conversation. To perform many units of business, the TPA may be instantiated as many long-running conversations (serially or concurrently) as is appropriate to the application and the processing capabilities of the parties' systems.

Figure 1 shows the main functions provided by the TPA. We now give a brief

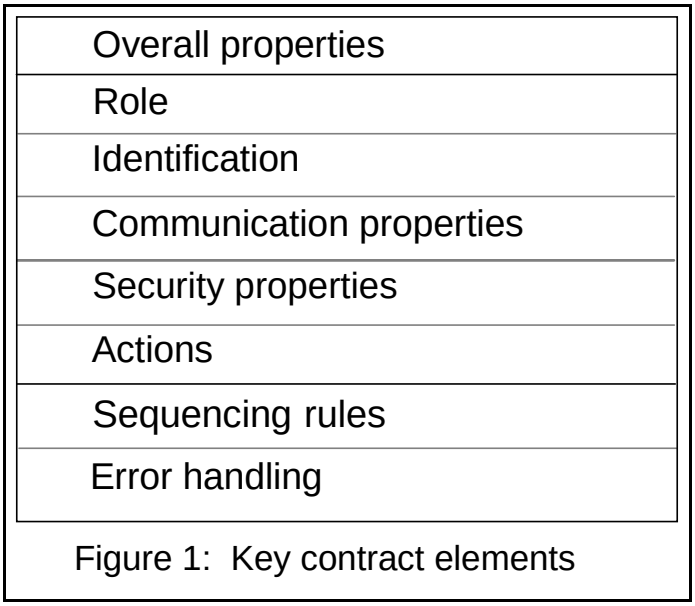



Figure 1: Key contract elements

overview of these functions. Section 4 describes the actual TPA language.

Overall properties of the TPA include its name, starting and ending dates, and similar global parameters. The role section provides the means to define a TPA in terms of generic roles such as airline and hotel and to produce a specific instance of the TPA by substituting specific parties for the role parameters. The identification section specifies the organization names of the parties and various contact information such as e-mail and postal service addresses. It also optionally specifies an outside arbitrator to be used for settling disputes. Communication and security properties include communication protocol (e.g. HTTP, SMTP), communication addresses, authentication and nonrepudiation protocols, certificate parameters, etc.

For each party which can act as a server, there is an action menu which lists the actions that the other party can request and various characteristics of those actions. Sequencing rules specify the order in which actions can be requested on each server. Error handling rules are various conditions related to error conditions, such as the maximum waiting time for the response to a request.

## 4. Business to Business TPA Language

The TPA is an XML document from which code is generated at each of the trading partners' computer systems. Authoring and code-generation tools are provided, as will be described later. The TPA document is described by an XML Document Type Definition (DTD) or XML-Schema file, which defines the tree structure of the TPA tags and some XML syntactic rules but not rules defining specific values of the tags or the semantic interrelations among the tags. These semantics are defined by a textual design document and are embodied in rules, understood by the authoring tool, which aid in the creation of a valid TPA.

### 4.1 Overall Structure

The overall XML structure of the TPA is as follows. Each of these tags is the top level of a subtree of tags (subelements). We will illustrate the following discussion with snippets of XML.

```
<TPA>
     <TPAInfo>  <!-- TPA preamble -->
          ...  <!--TPAname, role definitions,
               participants, etc.-->
          </TPAInfo>
     <Transport>
          ... <!--communication and transport
                   security information-->
          </Transport>
     <DocExchange>
          ... <!--document-exchange and message security
              information-->
          </Security>
     <BusinessProtocol>
          <ServiceInterface>  <!-- for each provider-->
               ...  <!--Action definitions etc.-->
               </ServiceInterface>
          </DocExchange>
     </TPA>
```

### 4.2 Layer Structure of TPA

The <BusinessProtocol>, <DocExchange>, and <Transport> sections describe the processing of a unit of business (conversation). These sections form a layered structure somewhat analogous to a layered communication model.

**Business-Protocol Layer:** The Business-Protocol layer defines the heart of the business agreement between the trading partners: the services (actions) which parties to the TPA can request of each other and sequencing rules that determine the order of requests. The Business-Protocol layer is the interface between the TPA-defined actions and the business-application functions that actually perform the actions.

**Document-Exchange layer**: The Document Exchange layer accepts a business document from the Business Protocol layer, optionally encrypts it, optionally adds a digital signature for nonrepudiation, and passes it to the transport layer for transmission to the other party.

**Transport layer**: The transport layer is responsible for message delivery using the selected communication protocol. Transport security (encryption and authentication) definitions are also provided.

### 4.3 Roles

When a given TPA can be repeatedly reused for different groups of parties, a prototype TPA or template can be written in terms of role parameters rather than specific party names. The authoring tool can then generate a specific TPA by substituting party names for the role parameters and filling in specifics of those parties such as their electronic addresses. The role definitions are included under the <TPAInfo> tag. Each <RoleDefn> tag supplies a pair of role parameter and actual name The <RoleName> tag defines the name of each role. The <RolePlayer> tag has a blank value in a TPA template and the name of an actual party in a specific TPA. Here is the XML for the role definitions for a TPA between an arbitrary airline (@airline) and an arbitrary hotel (@hotel). In this example, the tags under <Role> particularize the TPA to an agreement specifically between Hotelco and Airlineco.

```
<Role>
     <RoleDefn>     <!--one or more-->
          <RoleName>@hotel</RoleName>
          <RolePlayer>Hotelco</RolePlayer
          </RoleDefn>
     <RoleDefn>
          <RoleName>@airline</RoleName>
          <RolePlayer>Airlineco</RolePlayer>
          </RoleDefn>
     </Role>
```

When the authoring tool replaces the role parameters by actual party names, it either asks the author for party-specific information or finds this information in a previously-built database.

### 4.4 Transport Layer

In the transport layer, the communication properties section (<Communication> tag) defines the details of the system to system communication used in the application. These include the protocol to be used by both parties (e.g. HTTP, SMTP), each party's address parameters, maximum allowed network delay, and other parameters. Following is an example of the communication definition for HTTP:

```
<Communication>
   <HTTP>
     <Version>version</Version>
     <HTTPNode> <!--One for each party-->
          <OrgName Partyname=name/>
          <HTTPAddress>
               <URL URLName=type>url</URL>
               <!--additional URL tags as needed>
               </HTTPAddress>
          </HTTPNode>
     <NetworkDelay>time</NetworkDelay>  <!--Optional-->
     </HTTP>
   </Communication>
```

The transport-security properties tags (not shown) define the security protocols to be used in transporting messages. Protocols are defined for encryption and authentication. Encryption information includes the name of the encryption protocol and various parameters defining the certificates. Information supplied for authentication includes the type of authentication (e.g. password or certificate), the specific protocol (e.g. SSL), and the certificate parameters.

### 4.5 Document-Exchange Layer

Information included in the document-exchange layer includes the name of the protocol, such as OBI, the message-encoding choice (example: BASE64), whether or not duplicate messages should be detected, and the message-security definition. Message security may be either or both of digital-envelope (secret-key encryption using certificate-based encryption to exchange the secret keys) and certificate-based nonrepudiation.

### 4.6 Business-Protocol Layer

The `<BusinessProtocol>` tag defines the section of the TPA which contains all the business-protocol definitions that support the business application. Under `<BusinessProtocol>` is the service interface definition for each party that can act as a server. Each service interface contains some overall parameters and the action menu, which contains the set of definitions of the actions that this party will accept as service requests. The syntax is

```
<BusinessProtocol>
     <ServiceInterface>  <!--one or more-->
          ... <!-- action menu and other definitions-->
          </ServiceInterface>
     </BusinessProtocol>
```

### 4.7 Action Definition

For each party to the TPA which can act as a server, there is an action menu which identifies the permissible action requests and their characteristics. We discuss the main elements of an action definition using the following OBI buyer action definition (See "Application Example").

```
<Action>
   <Request>
     <RequestName>processOBIPOR</RequestName>
     <RequestMessage>OBIPOR</RequestMessage>
       <!--OBIPOR is a keyword which specifies the format of
       the message, in this case a purchase order request-->
    </Request>
   <Response>
       <ResponseName>handleOBIPO</ResponseName>
```

```
        <ResponseMessage>OBIPO</ResponseMessage>
        <ResponseServiceTime>
             <ServiceTime>3600</ServiceTime>
                  <!-- 1-hour maximum time -->
             </ResponseServiceTime>
        </Response>
    </Action>
```

The request name is `processOBIPOR`, i.e. the action transmits a purchase-order request to the OBI buyer. The `<Response>` tag indicates that the response is by means of an asynchronous message from the OBI seller server to the OBI buyer server and that the response causes the handleOBIPO method to be invoked at the issuer of the action (here, the OBI seller server). The response transmits a completed purchase order (`OBIPO`). The `<ResponseServiceTime>` tag specifies the worst case service time for the server (in this case, the OBI seller server) until the response is returned. Here, it is 3600 seconds, i.e. 1 hour. If the specified time is exceeded, it is up to the requester's application logic to decide what to do next.

Sequencing rules are used to specify the permissible order of action invocations on a given server. The permissible initial action or actions is specified as follows, specified under the `<ServiceInterface>` tag.

```
<StartEnabled>
     <RequestName>action_name</RequestName>
          <!--one for each action permitted as the initial
              action-->
     </StartEnabled>
```

There is one `<StartEnabled>` tag for each party which can act as a server. Only one of the actions whose names are specified under `<StartEnabled>` may be invoked as the first action in a given conversation on that server.

Within each action definition, a sequencing rule specifies which actions can no longer be invoked following the completion of the particular action, and which actions become permissible following the particular action. The specification is as follows:

```
<Sequencing>
     <Enable>   <!--actions permitted after this one-->
          <RequestName>name_of_action</RequestName>
               ...
          </Enable>
     <Disable>  <!--actions not permitted after this one-->
          <RequestName>name_of_action</RequestName>
               ...
          </Disable>
          </Sequencing>
```

The `<Enable>` tag specifies which actions are permissible following the action whose definition contains the `<Sequencing>` tag. The `<Disable>` tag specifies which actions are

no longer permitted after this action. We are investigating the possible need to extend the sequencing rules to cover sequencing of actions across multiple servers.

Many error conditions are handled in standard ways by the framework and their handling is not specified in the TPA. For example, the framework automatically retries for failures to receive transport-level acknowledgments. Some errors, such as sequencing errors, may be severe enough for the parties to invoke the arbitrator to determine whether a TPA violation occurred. Duplicate messages are most likely to arise during recovery, when incomplete actions are retried. The TPA can specify that if the recipient recognizes a duplicate message, the duplicate can be ignored. If the duplicate is a request message, the server can then re-send the response message.

## 5. TPA Authoring and Code Generation

In order to utilize an electronic TPA, the TPA must first be composed and agreed to by the parties. Then registration information must be extracted from the TPA and the necessary executable code generated. There are many possible designs for the tools. The design choices for the code generator and registration tool, in particular, depend on the specifics of the system in which they work. There can be no requirement that the same code generator and registration tool be used by both parties to the TPA. We here describe the tools we are developing as part of the Coyote project (Dan *et al* 1998). In our project, these tools are implemented in Java.

Because the TPA is a complex document and XML is not an intuitive language, an authoring tool is essential in preparing a TPA. Once the TPA is verified as valid and agreed to by both parties, it is passed to the TPA registration tool at each party's site. This tool extracts some of the content and stores the content in the registration database.

The business logic registration tool is used to associate actions which were specified in the TPA with business functions of which is a service provider, so that when the an action is requested of the service provider, the correct sequence of business functions is called.

The code generation tool uses information from the TPA and the registration database to convert a collection of templates into the executable file.

### 5.1 Authoring Tool

There are two parts to creating a TPA. They are creating models of the tags and authoring a specific TPA, guided by the models. The authoring tool provides a way for an expert to prepare a model from which a TPA can be constructed by someone with far less knowledge of the required semantics. The model contains the TPA semantic information needed to guide a user in creating a correct TPA.

The authoring tool starts with a DTD or XML Schema document, which provides the syntactic structure of the TPA. Then it constructs a model of a general TPA by asking the model maker to provide examples (semantics) of all parts of the TPA. Once a model is complete, it is available

to any author who, by answering a few specific questions, can create a very complex TPA with a high probability of success. Figure 2 illustrates the process of creating a model and a TPA.

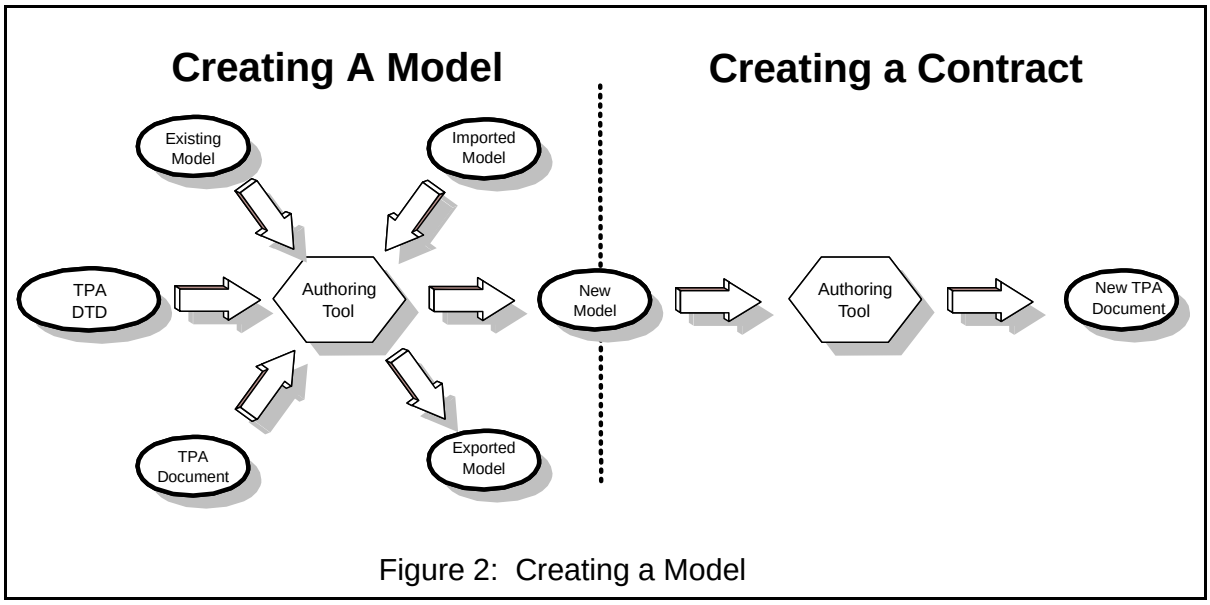



Figure 2: Creating a Model

A model consists of a collection of models of the tags to be used in the TPA. The models are in a tree structure which corresponds to the tree structure of the tags in the TPA. Each model of a tag is an example of the subtree under the tag. For example, a tag representing a communications protocol section has, as its subtree, information specific to a particular protocol.

The TPA author starts the authoring procedure after a model has been loaded. The authoring tool now uses the model to drive the authoring procedure. Starting with the root of the model, the authoring tool examines the choices for models beneath the root. If there is no choice to be made, the authoring tool accepts the model, proceeds to the next level, and repeats the above procedure for each child. If choices are to be made, a panel is displayed asking the user to select the correct model. The authoring tool then continues with that choice.

### 5.2 Code Generation

The code generator transforms the TPA into registration information and code which enforces the rules of interaction. A TPA object is created at the site of each party to the TPA. The code generation process is illustrated in Figure 3.

Code generation starts from a set of templates which consist of a combination of native (Java or any other) language and macro-style directives. These directives are written in a macro language consisting of information such as a basic set of data types, a basic set of functions used to obtain information from the TPA and other external sources, declaration statements, assignment

statements, and conditional statements which change the execution flow, depending upon values of variables and functions.

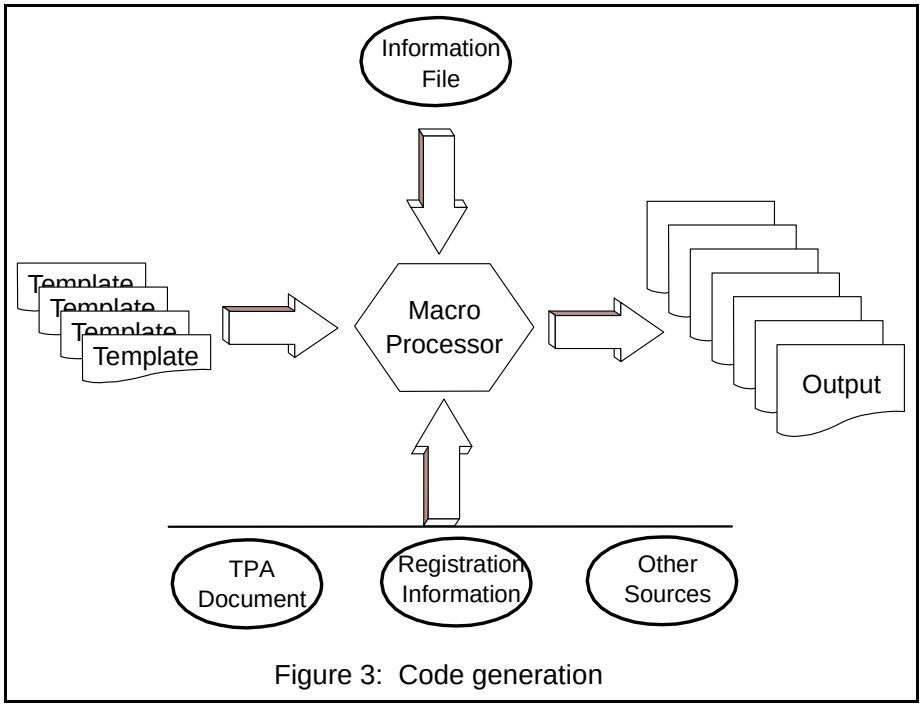



Figure 3: Code generation

A macro processor scans the template looking for directives. It executes any directives it encounters, and handles any native language statements as character strings, performing any needed processing, and writing the processed statements to a file.

## 6. Application Example

This section describes an example of the TPA and server structure. for an existing public protocol, OBI.

*Open Buying on the Internet (OBI),* Openbuying 1998, is a protocol for business-to-business Internet commerce. It was designed by the Internet Purchasing Roundtable and is supported by the OBI Consortium. OBI defines the procedures for the high-volume, low-dollar purchasing transactions that make up most of an organization's purchasing activity. In this section, we describe OBI, how it can be described by a TPA, and a schematic view of a possible implementation. Figure 4 illustrates the participants in an OBI transaction and the basic information flows. A complete OBI TPA is shown in the appendix.

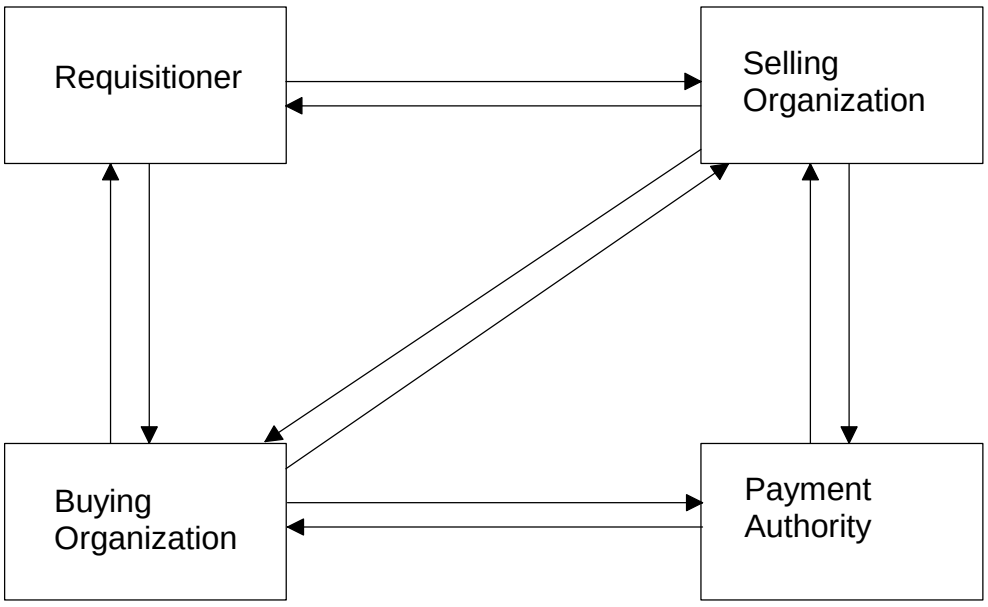



Figure 4: OBI Participants and Flows

The requisitioner is a member of the buying organization (e.g. an employee of a company) and is permitted to place orders directly with the selling organization's merchant server. The requisitioner can browse a catalog and place an order with the selling organization using a browser. When the requisitioner has placed an order, the selling organization's server sends a partial purchase order (purchase order request) to the buying organization's server. The buying organization validates the purchase order request and transforms it into a complete purchase order which it returns to the selling organization. The selling organization then prepares an invoice or otherwise arranges for payment and ships the ordered merchandise. The payment authority is an optional part of the system. Its purpose is to handle electronic payments. Using the browser, the requisitioner can also view and update various information at the buying organization server such as the requisitioner's profile, outstanding requests, etc. The requisitioner can also check the status of an order at the selling organization.

An additional possibility is that the buying organization can send an "unsolicited" purchase order to the selling organization without a prior request and partial purchase order initiated by a requisitioner. This mode might be used, for example, when a purchasing department purchases large volumes to supply a stock room.

In a one possible implementation of OBI, there is a TPA between the buying organization and the selling organization, each of which has a business to business server. In OBI terms, the TPA is a trading partner agreement (TPA). The payment authority, if present, is outside the scope of the

2-party TPA between buying organization and selling organization. It may interact with the buying organization and the selling organization in a variety of ways. The interaction may be through separate 2-party TPAs between the payment authority and the buyer and seller organizations. It may also be simply through application programs.

Following are the main functions included in the OBI TPA:

- Organization names of the parties to the TPA.
- Communication protocol definition. In this case it is HTTP, and includes the specific URLs of the buyer and seller.
- Security information such as the protocol (SSL in this case) and various certificate parameters
- Action menus for the buyer and the seller. The action list for the buyer is illustrated above in "Business to Business TPA Language". It consists of one action, "Process OBI Purchase Order Request". The completed purchase order is returned to the seller by means of a callback. The action list for the seller also consists of one action, "Process OBI Unsolicited Purchase Order".

Figure 5 shows the basic system structure and flow of an implementation of OBI. Shown in the figure are the TPA objects generated from the TPA at the buyer and seller servers. These objects provide the interfaces between various processes controlled by the TPA (in particular, the action requests) and the application logic at each server.

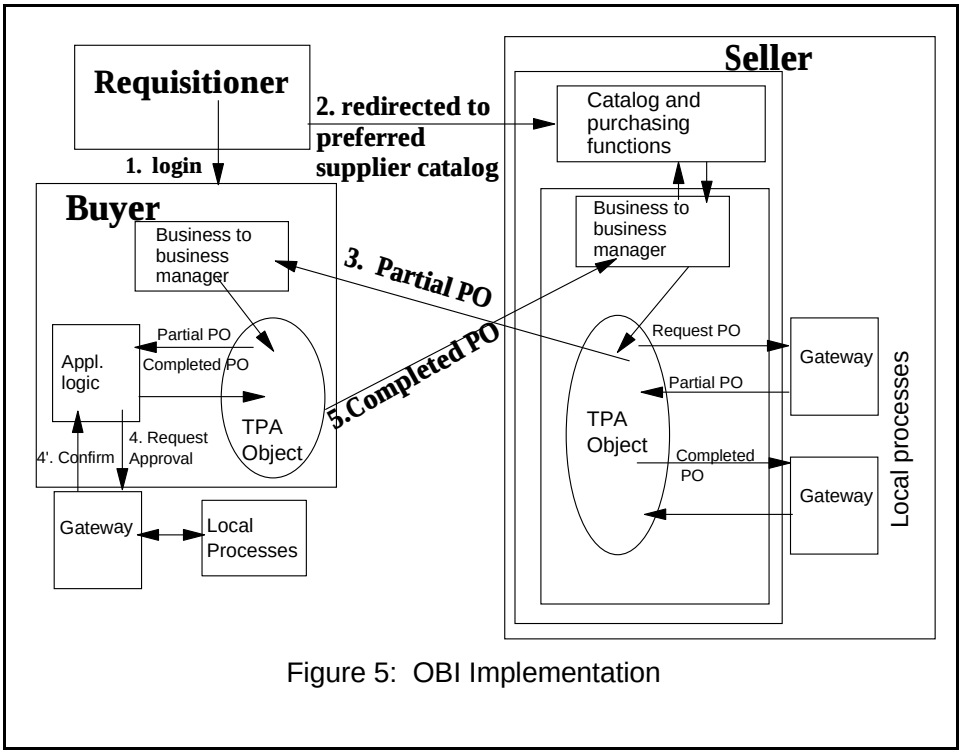



Figure 5: OBI Implementation

The process starts when a requisitioner contacts(1) the buyer server via a browser and is redirected(2) to the URL for the seller server. The requisitioner is shown the supplier catalog

appropriate to the requisitioner's organization. When the requisitioner makes a selection, the request is communicated to the TPA object. The TPA object communicates the purchase request to the local business processes via one of the gateways shown at the far right in the figure. A partial purchase order is returned to the TPA object via the gateway. The TPA object then issues the `processOBIPOR` action request(3) to the buyer server, sending a partial purchase order to the buyer server.

This request arrives at the buyer's TPA object, which evaluates the rules defined in the TPA and then sends the partial purchase order to the buyer application logic. In processing the partial purchase order, the application logic communicates with local business processes, via the gateway shown at the lower left in the figure, to request approval(4) of the purchase order. If the purchase is approved(4'), the approval arrives at the application logic, which completes the purchase order and passes the completed purchase order to the buyer's TPA object. The TPA object then issues the callback(5), sending the completed purchase order back to the seller.

The completed purchase order arrives at the seller's TPA object, which passes it to the local processes via the gateway at the lower right. The local processes handle fulfillment (e.g. shipping) and invoicing/payment. They also initiate a confirmation message to be returned to the requisitioner via the browser (not shown in the figure).

## 7. Future Work

We are extending the TPA ideas and language to areas such as TPA hierarchy, linking of multiple TPAs, and dynamic negotiation. We are also investigating TPAs in which there are more than two parties.

In addition, we are investigating how to incorporate business constraints into the TPA. Business constraints are conditions placed on data items in response messages. The results of these tests may modify further processing within the TPA. An example is a test of whether a cancellation action (e.g. to cancel a reservation) was issued during the allowed time range after the original action.

## 8. Summary

This paper has discussed various issues in inter-business electronic interactions and in the use of an electronic TPA for embodying the IT-related and business protocol terms and conditions used in business to business electronic commerce. We have designed an XML-based TPA language and tools for authoring TPAs in that language and generating code from the TPAs. We described examples of two applications which make use of TPAs and showed schematic views of such systems.

## Acknowledgments

The authors express their appreciation to the following for contributions to the formulation of the TPA principles and language: Francis Parr, Vibby Gottemukkala, Terry Lau, Satwinder Brar, George Kleon, Gerald Anderson, John Ibbotson, Christine Draper, Linh Lam, Stewart Palmer, Richard King, and Sastry Duri.

## References

Corba: *The Common Object Request Broker Architecture and Specification,* Rev. 2.2, Object Management Group, http://www.omg.org, 1998.

Dan, A., Dias, D., Nguyen, T., Sachs, M., Shaikh, H., King, R., and Duri, S., The coyote project: framework for multi-party e-commerce, *Proc. Research and Advanced Technology for Digital Libraries, Second European Conference,* ECDL'98, Heraklion, Greece, Sept. 1998, Springer Verlag, Berlin, Germany, 1998, p. 873-889.

Dan, A. and Parr, F. An object implementation of network centric business service applications (NCBAs), *OOPSLA Business Object Workshop,* Atlanta, GA, USA, Sept. 1997a.

Dan, A. and Parr, F., The coyote approach for network centric business service applications, *HPTS Workshop,* Asilomar, CA, USA, 1997b.

Dan, A. and Parr, F. Long running application models and cooperating monitors, submitted to *HPTS workshop,* Asilomar, CA, 1999.

Ejb: *Enterprise Java Beans Specification,* ver. 1.1, http://www.javasoft.com/products/ejb, 1999.

Garcia-Molina, H. and Salem, K., SAGAS, *Proc. of ACM SIGMOD Conf.,* Association for Computing Machinery, New York, NY,1987, pp. 249-259.

Gray, J. and Reuter, A., *Transaction Processing: Concepts and Techniques,* Morgan-Kaufmann, San Mateo, CA, 1993.

Openbuying: *Open Buying on the Internet Technical Specifications,* Release V1.1, The Open Buying on the Internet (OBI) Consortium, http://www.openbuy.org, 1998.

Weigand, H. and Ngu, A., Flexible specification of interoperable transactions, *Data & Knowledge Engineering,* Vol. 25, 1998, pp. 327-345.

Wfmc: *The Workflow Management Coalition Specification,* http://www.wfmc.org, 1998.

## Appendix: OBI TPA

Following is a TPA which defines OBI.

```
<?xml version="1.0"?>
<!DOCTYPE TPA SYSTEM "TPA.xsd" >
<!--************************************************************************-->
<!--              OBI TPA      between Large Co (buying company)           -->
<!--                     and Pens Are We (selling company)                  -->
<!-- (C) Copyright IBM Corporation 2000                                     -->
<!--************************************************************************-->
<TPA xmlns="tpa.xsd">
<!--************************************************************************-->
<!-- General information                                                    -->
<!--************************************************************************-->
  <TPAInfo>
    <TPAName>OBIStandard</TPAName>
  <TPAType>
    <Protocol>OBI</Protocol>
    <Version>1.0</Version>
    <Type>SS</Type>
   </TPAType>
<!--************************************************************************-->
    <Participants>
<!--************************************************************************-->
<!-- Specification of Buyer                                                 -->
<!--************************************************************************-->
      <Member IdCodeType="ZZ" MemberId="777777777777777">
        <PartyName Partyname="_LargeCo">Large Co</PartyName>
        <CompanyTelephone>914-945-3000</CompanyTelephone>
        <Address>
          <AddressType>location</AddressType>
          <AddressLine>Large Co</AddressLine>
          <AddressLine>HQ Building</AddressLine>
          <AddressLine>1 Main Street</AddressLine>
          <City>SmallTown</City>
          <State>NY</State>
          <Zip>10000</Zip>
          <Country>USA</Country>
        </Address>
        <Address>
          <AddressType>billing</AddressType>
          <AddressLine>Large Co</AddressLine>
          <AddressLine>Accounting Department</AddressLine>
          <AddressLine>100 Bean Counters Road</AddressLine>
          <City>Any City</City>
          <State>CT</State>
          <Zip>06000</Zip>
          <Country>USA</Country>
        </Address>
        <Address>
          <AddressType>shipping</AddressType>
          <AddressLine>Large Co</AddressLine>
          <AddressLine>Procurement Department</AddressLine>
          <AddressLine>99 Purchase Road</AddressLine>
          <City>Buy City</City>
          <State>NY</State>
          <Zip>10001</Zip>
          <Country>USA</Country>
        </Address>
        <Contact Type = "primary">
          <LastName>Smith</LastName>
          <FirstName>John</FirstName>
```

```
          <MiddleName>L.</MiddleName>
          <Title>Senior Buyer</Title>
          <ContactTelephone Type = "primary">914-111-6789
</ContactTelephone>
          <ContactTelephone Type = "secondary">914-111-6790
</ContactTelephone>
          <EMail Type = "primary">jjsmith@largeco.com</EMail>
          <EMail Type = "secondary">
            http://www.largeco.com/procurement/jsmith.html
            </EMail>
          <Fax>914-111-6780</Fax>
        </Contact>
        <Contact Type = "secondary">
          <LastName>Blow</LastName>
          <FirstName>Joe</FirstName>
          <MiddleName>J.</MiddleName>
          <Title>Buyer</Title>
          <ContactTelephone Type = "primary">914-111-6722
</ContactTelephone>
          <ContactTelephone Type = "secondary">914-111-6725
</ContactTelephone>
          <EMail Type = "primary">jblow@largeco.com</EMail>
          <Fax>914-111-6780</Fax>
        </Contact>
      </Member>
<!--*********************************************************************-->
<!-- Specification of Seller                                             -->
<!--*********************************************************************-->
      <Member IdCodeType="ZZ" MemberId="888000009000000">
        <PartyName Partyname="_PensAreWe">Pens Are We
</PartyName>
        <CompanyTelephone>945-123-1000</CompanyTelephone>
        <Address>
          <AddressType>location</AddressType>
          <AddressLine>Pens Are We</AddressLine>
          <AddressLine>Building 001</AddressLine>
          <AddressLine>123 High Street</AddressLine>
          <City>EarthQuake City</City>
          <State>CA</State>
          <Zip>94567</Zip>
          <Country>USA</Country>
        </Address>
        <Contact Type = "primary">
          <LastName>Doe</LastName>
          <FirstName>Jane</FirstName>
          <MiddleName>E.</MiddleName>
          <Title>Vice President of Internet Sales</Title>
          <ContactTelephone Type = "primary">945-123-4567
</ContactTelephone>
          <ContactTelephone Type = "secondary">945-123-4570
</ContactTelephone>
          <EMail Type = "primary">janedoe@pensarewe.com</EMail>
          <EMail Type = "secondary">
            http://www.pensarewe.com/sales/jdoe.html
            </EMail>
          <Fax>945-123-9999</Fax>
        </Contact>
      </Member>
<!--*********************************************************************-->
<!-- Specification of Arbitrator                                         -->
<!--*********************************************************************-->
      <Arbitrator IdCodeType="01" MemberId="888000009000001">
        <PartyName Partyname="_XYZArbitrator">XYZArbitrator</PartyName>
        <CompanyTelephone>780-333-1111</CompanyTelephone>
        <Address>
```

```
          <AddressType>location</AddressType>
          <AddressLine>XYZArbitrator</AddressLine>
          <AddressLine>Suite 3</AddressLine>
          <AddressLine>77 Lawyers Blvd</AddressLine>
          <City>ABC City</City>
          <State>MA</State>
          <Zip>01234</Zip>
          <Country>USA</Country>
        </Address>
        <Contact Type = "primary">
          <LastName>Black</LastName>
          <FirstName>Joe</FirstName>
          <MiddleName>K.</MiddleName>
          <Title>Mr.</Title>
          <ContactTelephone Type = "primary">780-333-4040
</ContactTelephone>
          <ContactTelephone Type = "secondary">780-333-4045
</ContactTelephone>
          <EMail Type = "primary">jblack@xyzarbitrator.com</EMail>
          <EMail Type = "secondary">
            http://www.xyzarbitrator.com/jblack.html</EMail>
          <Fax>780-333-5000</Fax>
        </Contact>
      </Arbitrator>
    </Participants>
    <Duration>
      <Start>
        <Date>01/01/1999</Date>
        <Time>00:00:00</Time>
      </Start>
      <End>
        <Date>01/01/2001</Date>
        <Time>00:00:00</Time>
      </End>
    </Duration>
    <InvocationLimit>100000</InvocationLimit>
    <ConcurrentConversations>1</ConcurrentConversations>
    <ConversationLife>86400</ConversationLife>
  </TPAInfo>
<!--************************************************************************-->
<!-- Specification of Transport Protocol #01                                -->
<!--************************************************************************-->
  <Transport>
   <Communication>
    <HTTP>
     <HTTPNode>
      <OrgName Partyname="_LargeCo"/>
      <HTTPAddress>
        <URL URLName="requestURL">
            https://www.largeco.com/jackal/servlet/OBIBuy</URL>
      </HTTPAddress>
     </HTTPNode>
     <HTTPNode>
      <OrgName Partyname="_PensAreWe"/>
      <HTTPAddress>
        <URL URLName="logOnURL">
            https://www.pensarewe.com/coyote/servlet/OBILogon</URL>
        <URL URLName="requestURL">
            https://www.pensarewe.com/coyote/servlet/OBIsell</URL>
        <URL URLName="responseURL">
            https://www.pensarewe.com/coyote/servlet/OBIsell</URL>
      </HTTPAddress>
     </HTTPNode>
    <NetworkDelay>300</NetworkDelay>
    </HTTP>
```

```
   </Communication>
<!--*************************************************************************-->
<!-- Specification of Transport Security Protocol                            -->
<!--*************************************************************************-->
   <TransportSecurity>
      <Encryption>
        <Protocol>SSL</Protocol>
        <Version>3.0</Version>
        <Certificate>
          <CertType>X509.V3</CertType>
          <KeyLength>1024</KeyLength>
          <Party>
            <OrgName Partyname="_LargeCo"/>
            <IssuerOrgName>VeriSign, Inc.</IssuerOrgName>
            <IssuerCertSource>http://www.verisign.com/certs
</IssuerCertSource>
          </Party>
          <Party>
            <OrgName Partyname="_PensAreWe"/>
            <IssuerOrgName>GTE, Inc.</IssuerOrgName>
            <IssuerCertSource>http://www.gte.com/certs
</IssuerCertSource>
          </Party>
        </Certificate>
      </Encryption>
      <Authentication>
        <CertificateAuthen>
        <Protocol>SSL</Protocol>
        <Version>3.0</Version>
        <Certificate>
          <CertType>X509.V3</CertType>
          <KeyLength>1024</KeyLength>
          <Party>
            <OrgName Partyname="_LargeCo"/>
            <IssuerOrgName>VeriSign, Inc.</IssuerOrgName>
            <IssuerCertSource>http://www.verisign.com/certs
</IssuerCertSource>
          </Party>
          <Party>
            <OrgName Partyname="_PensAreWe"/>
            <IssuerOrgName>GTE, Inc.</IssuerOrgName>
            <IssuerCertSource>http://www.gte.com/certs
</IssuerCertSource>
          </Party>
        </Certificate>
       </CertificateAuthen>
      </Authentication>
     </TransportSecurity>
    </Transport>
<!--*************************************************************************-->
<!-- Specification of DocExchange Protocol                                   -->
<!--*************************************************************************-->
  <DocExchange>
    <DocExchangeProtocol>OBI</DocExchangeProtocol>
    <MessageEncoding>BASE64</MessageEncoding>
    <MessageIdempotency>yes</MessageIdempotency>
<!--*************************************************************************-->
<!-- Specification of Message Security                                       -->
<!--*************************************************************************-->
    <MessageSecurity>
      <NonRepudiation>
        <Protocol>DigitalSignature</Protocol>
        <HashFunction>MD5</HashFunction>
        <EncryptionAlgorithm>RSA</EncryptionAlgorithm>
        <SignatureAlgorithm>DSA</SignatureAlgorithm>
```

```
        <Certificate>
          <CertType>X509.V3</CertType>
          <KeyLength>1024</KeyLength>
          <Party>
            <OrgName Partyname="_LargeCo"/>
            <IssuerOrgName>Verisign Inc.</IssuerOrgName>
            <IssuerCertSource>http://www.verisign.com/certs
               </IssuerCertSource>
          </Party>
          <Party>
            <OrgName Partyname="_PensAreWe"/>
            <IssuerOrgName>GTE Inc.</IssuerOrgName>
            <IssuerCertSource>http://www.gte.com/certs</IssuerCertSource>
          </Party>
        </Certificate>
      </NonRepudiation>
    </MessageSecurity>
  </DocExchange>
  <BusinessProtocol>
<!--************************************************************************-->
<!-- Specification of Service Interface 01                                  -->
<!--************************************************************************-->
   <ServiceInterface InterfaceId="interface01">
    <OrgName Partyname="_LargeCo"/>
    <Client>
      <OrgName Partyname="_PensAreWe"/>
    </Client>
    <ActionMenu>
      <Action ActionId="action01" Type="basic">
        <Request>
          <RequestName>putOPOR</RequestName>
          <RequestMessage>OBIPOR</RequestMessage>
        </Request>
        <Response>
          <ResponseName>getOPO</ResponseName>
          <ResponseMessage>OBIPO</ResponseMessage>
          <ResponseServiceTime>
            <ServiceTime>3600</ServiceTime>
            <Presume>fail</Presume>
          </ResponseServiceTime>
        </Response>
      </Action>
    </ActionMenu>
    <ServerServiceTime>
        <ServiceTime>3660</ServiceTime>
        <Presume>fail</Presume>
        </ServerServiceTime>
    <StartEnabled>
      <RequestName>putOPOR</RequestName>
    </StartEnabled>
   </ServiceInterface>
<!--************************************************************************-->
<!-- Specification of Service Interface 02                                  -->
<!-- This interface below is for UnSolicited OBIPO from                     -->
<!-- buying organization to selling organization                            -->
<!--************************************************************************-->
  <ServiceInterface InterfaceId="interface02">
    <OrgName Partyname="_PensAreWe"/>
    <Client>
      <OrgName Partyname="_LargeCo"/>
    </Client>
    <ActionMenu>
      <Action ActionId="action03" Type="basic">
        <Request>
          <RequestName>shop</RequestName>
```

```
            <!--Initiates shopping at merchant server-->
            <RequestMessage>shopMessage</RequestMessage>
          </Request>
        </Action>
        <Action ActionId="action02" Type="basic">
          <Request>
            <RequestName>putOPO</RequestName>
            <RequestMessage>OBIPO</RequestMessage>
          </Request>
        </Action>
      </ActionMenu>
      <ServerServiceTime>
          <ServiceTime>3660</ServiceTime>
          <Presume>fail</Presume>
      </ServerServiceTime>
    </ServiceInterface>
  </BusinessProtocol>
</TPA>
```

# Implementing an Industry e-Business Initiative: Getting to RosettaNet

Patricia J. O'Sullivan, IT Strategy & Technology, Intel Corp.
Don S. Whitecar, PE, IT e-Business Integration, Intel Corp.

Index words: e-Business, e-Commerce, standards, standards implementation, enterprise application integration, EAI, RosettaNet, B2B, trading partner automation, trading partner integration, B2B gateway, XML, WWW

## ABSTRACT

As Intel looked at the cost of its own successful early implementation of Web-based e-Commerce, it became clear that an industry-wide standards-based approach to e-Business is the only way to go.

We decided to help build the right business-to-business (B2B) specifications with the right industry initiative (RosettaNet* is our main focus) and then implement those specifications. As early adopters, this has turned out to be much more of an enterprise readiness effort than initially appreciated. Team composition, technical and business knowledge coalescence, formal and informal communication channels, cross-enterprise visibility, and establishment of appropriate resource levels are just some of the challenges we face.

We anticipate that an evolving, more robust infrastructure, together with lessons learnt from pilot projects, team experience, and more mature standards will lead to the full realization of expected benefits from RosettaNet. However, we offer here a "readiness model" that we hope can be used by others to "spin up" faster.

## INTRODUCTION

In early 1998, Paul Otellini, then Sr. Vice President of Intel's Sales and Marketing Group, crystallized much of our early thinking and experimenting with Internet-based e-Commerce into a simple challenge: take in $1Billion in sales orders via the Web in Q4'98. We took our first such order in July 1998 and had arrived at $1B *per month* by the start of Q4'98—success beyond our wildest dreams!

So, with that success, one may well ask what the problem is. Well, each customer's internal processes and systems are almost always different from ours, so we had no way to ensure that our applications, which worked well for us, did not introduce extra work or become otherwise burdensome for our customers. And, our customers buy many products from many suppliers in order to make up their complete product lines, so they are potentially facing extra work from *each* of their suppliers. As we moved forward with various plans to Internet-enable the way we do business with our customers, we also realized that the part of Intel that *buys* products was getting ready to establish a whole series of web-based procurement applications that we wanted our suppliers to use.

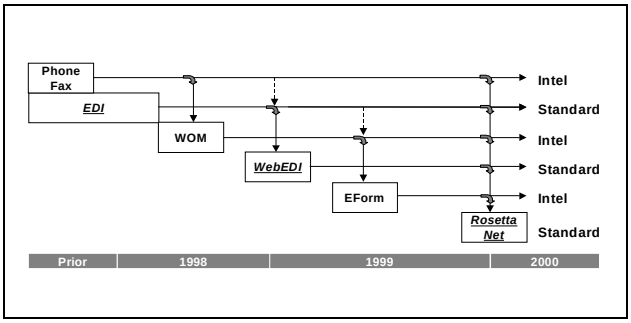
So, not only were we building a suite of applications that did not necessarily optimize e-Business for our trading partners (the phrase used generally in the e-Business arena to refer to other companies with which we do business, whether as customers, suppliers, or other), but we were facing the prospect of developing and/or buying a whole slew of applications that we (and our trading partners) would have to support and maintain over time.

What we needed were standards! However, we did **not** want standards that took years to develop; rather, we needed those that evolved at the same pace as the Internet, at the same pace as the emerging "killer app" of e-Business, and at the same pace as the technologies that underlay that growth. Furthermore, these standards had to focus on the real-world business processes of the supply chains that we are a part of, not those that tried to create a single universal e-Business solution or that sacrificed implementation to elegant technical solutions.

* Third-party brands and names are the property of their respective owners.

Of equal importance, these standards would need a sound, extensible architecture; would have to be adopted rapidly; and would have to be *demanded* by management and supported by business and technology stakeholders within our business environment.

Figure 1 illustrates the tension between some of Intel's e-Commerce solutions and our desire to use standard




solutions, as projected over time.

**Figure 1: Intel standards challenge**

As is true throughout history, but even more noticeable in today's Internet economy, timing is everything. In late 1997, an executive from one of the largest computing products distributors in the world approached us (as well as other leading players in the "IT Products" supply chain) with a vision and a plan. The vision was to create a common vocabulary and process set for e-Business in the context of a well understood supply chain. The plan was to pull together a fast-moving business consortium of companies representing over half the revenue of that supply chain, managed by a board of top executives from member companies, who would *precommit* to implementing the specifications that their members would jointly develop and vote upon.

At Intel, we pulled together a quick evaluation team, surveyed both our internal e-Business initiatives (such as Web Order Management (WOM), Supply Line Management (SLM), and eFORM) and the external e-Business standards/initiatives environment, and we quickly concluded that we needed to help realize this effort, and the sooner, the better. In the words of Colin Evans, Intel's Sales & Marketing e-Business architect, our competitive advantage would have to move from "first to move" to "first to standards."

And thus begins the lessons we have learned (and are still learning) about implementing business-to-business standards across the enterprise.

## INTEL'S ROSETTANET IMPLEMENTATION

Today we are preparing to meet our first major commitment to have a production-level implementation of at least one of the RosettaNet "partner interface process" (PIP) specifications (which are described more fully below) running on a robust infrastructure, with at least one trading partner. Among RosettaNet members, this milestone is known as "2.2.2000," which is the date that we will all be ready to demonstrate our success. We are in the thick of this implementation, learning lessons every hour.

The main focus of this paper is on what it takes to be internally *ready* to adopt and implement the suite of specifications collectively known as RosettaNet. We try to make these observations as concrete as possible, without being necessarily RosettaNet-specific. They should be of interest to anyone who is preparing to implement any standards-based approach to e-Business.

In order to understand the magnitude of our implementation effort (both initial and longer-term), it is necessary to provide a little background on the RosettaNet business and technical architecture. The bulk of this paper focuses on our use of a readiness model to ensure that our solutions could be deployed.

## ROSETTANET OVERVIEW

Although RosettaNet's supply chain scope began with IT products (e.g., boards, systems, peripherals, finished systems), it has expanded to include electronic components (e.g., chips, connectors). Intel obviously plays a role in both of these supply chains (often abbreviated as IT and EC). As maturity is gained in these environments, it is likely that RosettaNet's business scope will expand to other supply chains as well. Each supply chain's standardization efforts are overseen by a managing board composed of member company executives, who prioritize efforts, ensure synergy between supply chains as much as possible, and oversee resource allocation as administered by a paid staff.

RosettaNet focuses on three key areas requiring standardization in order to automate business interchanges between trading partners. First, vocabulary needs to be aligned; this includes both business and technical terminology germane to the transaction at hand. The RosettaNet Dictionary, drawing upon existing industry standards wherever possible, fills this need. Second, the way in which business messages are wrapped and transported must be specified. The RosettaNet Implementation Framework, which specifies the use of XML (Extensible Markup

Language), the World Wide Web (WWW), and other protocols serves this need. And third (and most important) the business processes governing the interchange of the business messages themselves must be analyzed, harmonized, and specified. RosettaNet terms these "Partner Interface Processes" or PIPs. Figure 2 shows these RosettaNet "ingredients."

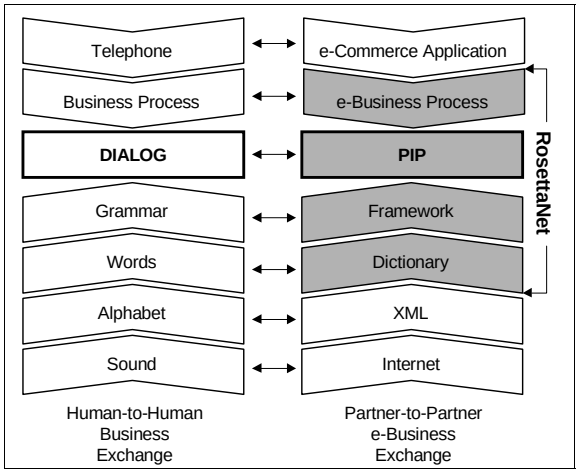



**Figure 2: RosettaNet ingredients**

To perform the work of analyzing, recommending, and documenting proposals for voting by the membership, RosettaNet member companies volunteer expert resources, both business and technical people, to lead and/or be a part of project teams. These people are either on part-time project duty or on detached full-time (short-term) assignments.

At present, six "clusters" of business activities (such as "Order Management") have been identified as initial targets of RosettaNet standardization efforts by the RosettaNet Managing Boards. Within those clusters, "segments" have been identified (e.g., within the Order Management cluster, "Quote & Order Entry" is one of four segments. Each segment is then analyzed in workshops that identify the necessary PIPs and document the choreography and business requirements around each PIP. RosettaNet anticipates that between 100 and 120 PIPs will result from the six clusters.

It is worth noting that when the second supply chain (EC) was added to RosettaNet's business scope, only two additional segments (and no additional clusters) had to be added. There is more synergy among related supply chains than many had guessed; this gives us more optimistic expectations for the addition of related supply chains.

Current RosettaNet clusters and segments are as follows:

- review segments: partner review; product/service review
- product introduction segments: preparation for distribution; product change notification
- marketing management segments: marketing campaign management; lead and opportunity management; design win management (EC only)
- order management segments: quote and order entry; transportation and distribution; product configuration; returns and finance management
- inventory management segments: price protection; collaborative forecasting; inventory allocation and replenishment; inventory and sales reporting; ship from stock and debit/credit (EC only)
- service and support segments: warranty management; asset management; technical support

RosettaNet member companies are increasingly realizing that, although the PIPs specify processes only at the point of interface between trading partners, the full value of their implementations will come when they align their internal processes with the PIPs as well. This makes it all the more imperative to have a tool with which to evaluate internal readiness for making the shift to standards-based e-Business.

## IMPLEMENTATION READINESS MODEL

Implementing a business-to-business (B2B) message exchange environment such as RosettaNet[*] has turned out to be much more of an enterprise readiness effort than initially anticipated. Team composition (size, diversity), technical and business knowledge coalescence, communication channels, at-large evangelism, cross-enterprise visibility, and establishing appropriate resource levels are just some of the challenges.

As an early adopter, we of course experience more pain than those who will follow. More supply chain experience, B2B gateway products, internal infrastructure, pilot projects, internal experience, and standards maturation will all ease the way. However, knowing where to look for "readiness" (or lack thereof) is critical to putting together a workable implementation

[*] Third-party brands and names are the property of their respective owners.

plan. To that end, our Implementation Readiness Model has identified four primary and six secondary readiness tracks to date.

The primary readiness tracks are as follows:

1. business strategy
2. B2B infrastructure
3. business process
4. application development

The secondary readiness tracks are as follows:

1. B2B external initiative
2. trading partner
3. solution provider
4. legal
5. security
6. audit

These tracks were identified as Intel went through the following process:

- early pilot (proof of concept) initiated by our Sales and Marketing Group (Internet Marketing and e-Commerce organization) and one trading partner (completed in August 1998)
- engagement with Intel IT (e-Business Integration) to provide necessary infrastructure to ramp into production mode
- involvement of Intel Planning and Logistics Group to rationalize current business processes and ERP (Enterprise Resource Planning) systems with RosettaNet processes
- cross-enterprise collaboration to get to 2.2.2000

Our next steps will be to drive a significant increase in participation by business process and application development groups in order to further deploy RosettaNet.

We are using this Readiness Model to help us get there.

## BUSINESS STRATEGY READINESS

The purpose of this track is to assess the maturity of an enterprise's B2B strategy at large. This is important because B2B solutions are currently strategically divided between browser-based (user-interface) on-demand applications and automated service applications that do not require user interfaces. RosettaNet implementers are primarily focused on trading partner automation to either rehost their Electronic Data Interchange (EDI) processes or reduce the need for browser-based applications. At present, little or no attention is being focused on feeding a B2B gateway with RosettaNet messages generated by browser-based application backend processes. Implementers should recognize that over time, as more trading partner business processes are automated, RosettaNet would reduce the need for on-demand B2B applications that provide a user interface. Therefore, many current browser-based B2B projects, funding initiatives, and roadmaps need to be re-evaluated to see whether an end of life timeframe exists. Key criteria in this track include a company's B2B strategy and a company's buy-side and sell-side motivations.

Intel, like many other companies, is using the Internet as a means to improve and simplify processes and services with its trading partners. Due to RosettaNet and similar initiatives, B2B solutions are evolving from trading partner portals or point applications requiring user interaction to automated solutions. This B2B automation evolution is enabling a shift from "engage customer eyeballs" to "customer at work," allowing customers (indeed, all trading partners) to use their own internal solutions while having immediate access to all information within their global enterprise. In other words, by enabling RosettaNet, companies should be able to reduce overall data entry and data interpretation costs. Automated data exchange and processes provide for higher quality data and faster processing time, as well as create the possibility of an event-driven global enterprise.

The motivations to implement RosettaNet may be greater within a company's buy-side or sell-side. However, implementing RosettaNet ultimately needs to encompass both the buy-side and sell-side of an enterprise's at-large B2B strategy. Implementers need to be sure to identify benefits by looking at the entire supply chain; that is, their customers' customers through their suppliers' suppliers. (This level of impact upon one's strategy will depend greatly upon how much of a company's purchased materials and finished products fall within the RosettaNet consortium scope of coverage.) Implementers should identify their other supply chains on both their sell and buy sides, and they should investigate how other B2B initiatives for trading partner automation are evolving.

The push to implement RosettaNet currently appears to be driven more by buyers in an attempt to simplify and improve productivity and margins. This may be because buyers naturally tend to engage suppliers with whom they can work more easily. However, as B2B automation spreads, we should see suppliers using their proven benefits to persuade their non-automated

customers to participate in automated services. For example, a seller should be able to provide its customers with improved pricing and availability when its customers provide real-time demand and inventory/sales out reporting instead of infrequent non-automated inputs. (Now, imagine the gains if an entire supply chain were to automate its demand and inventory/sales out reporting from end to end—this is one of RosettaNet's objectives).

A company's B2B strategy also needs to take into account the integration of mergers and acquisitions. Another of RosettaNet's benefits would be improved flexibility and agility as companies grow their core business by enabling a standard message exchange framework.

Finally, a company's B2B strategy needs to recognize the full potential of RosettaNet. Through the use of a self-describing message structure that includes a supply-chain dictionary-driven schema and meta-model for more than 100 business processes, RosettaNet supplies a strategic benefit. This message structure holds the potential of becoming a de facto message exchange standard in the near future as agent-to-service and service-to-service architectures evolve. The RosettaNet message structure is in fact sufficiently rich that it could be said to be a document database; RosettaNet messages could be used as disconnected documents passed between applications and databases within a disparate and distributed architecture.

### B2B Infrastructure Readiness

The purpose of this track is to define the infrastructure and to assess the level of effort needed to achieve it. This is important because becoming RosettaNet-compliant is only a small portion of the big picture. Although RosettaNet specifies a message structure, a message dictionary, a message exchange framework, and a message exchange protocol, it does not specify the infrastructure needed nor the backend processes required to receive, process, or send messages. Infrastructure is individually managed by each trading partner. Key criteria in this track include infrastructure components, e-Business standards and guidelines, and B2B gateway capabilities.

B2B message integration involves both public and private aspects. Receiving, unpacking, and routing a message, or assembling and sending a message (the public part) is relatively easy. The private (and more difficult) part of message integration includes process automation, workflow, and application integration that link into enterprise applications. In other words, the private part is the intra-enterprise application integration portion of enterprise application integration (EAI), while RosettaNet is the public trading partner application integration part of EAI.

RosettaNet is targeted for use within e-Business applications, predominantly B2B service-to-service applications; however, much of the infrastructure needed to support this is the same as required for B2C (business-to-consumer) and B2B browser-based applications. Many infrastructure components need to exist in order to proceed. Major infrastructure elements include an e-Business "landing zone," facilities, firewalls, proxy servers, networks, routers, communication services, and web servers.

A B2B gateway is needed. It must provide for inbound message receipt, authentication, authorization, entitlement, logging, and routing to a process automation workflow tool. This gateway also must support outbound message construction, packaging and logging. The B2B gateway must be able to provide RosettaNet-compliant messaging and also should be capable of supporting other B2B specifications, perhaps even EDI, file transfer protocol (FTP) and simple mail transfer protocol (SMTP) transport protocols. The build vs. buy study needs to be completed, while looking at maturing product offerings from solution providers.

RosettaNet provides for two basic types of messages: a transaction process message (Figure 3); and a subscription model message (Figure 4).

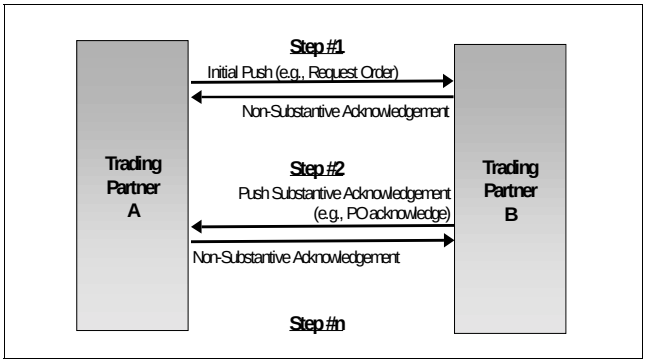



**Figure 3: Transaction process model**

In order to implement subscriptions, a collection of document repository, subscription, notification, and publication services needs to be provided. This is potentially a very large effort, and again, a few solution providers are working in this space.

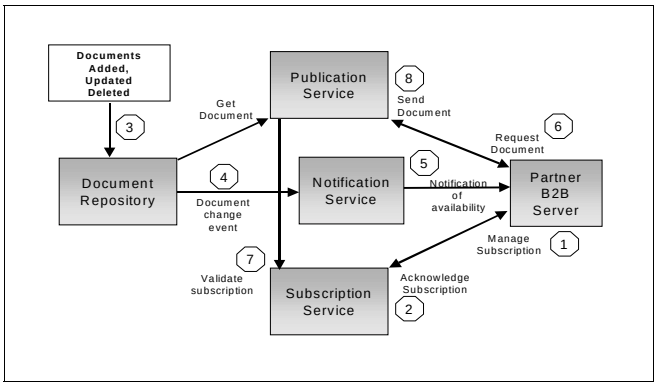



**Figure 4: Subscription model**

The major services within a B2B gateway include the following:

1. a trading partner database for a directory of trading partners, trading partner processes, and process parameters and their entitlements
2. non-repudiation (legal proof) archiving of message origin and content
3. public key infrastructure (PKI) repository of digital certificates and signatures for encryption, authorization, and authentication
4. PIP templates for integration to public and private process/workflow automation processes
5. virus detection capabilities for message attachments

The B2B gateway will likely coincide or integrate with existing gateways for FTP, value added network (VAN/EDI), and SMTP. Each of these gateways should comply with similar guidelines, designs, and implementations of authorization, entitlement, authentication, privacy, confidential document, and legal trade agreement practices.

RosettaNet is based upon the hypertext transfer protocol/secure (HTTP/HTTPS) protocol in an automated service-to-service framework that does not need visible or attended Web pages. Because HTTPS is needed for security, internal corporate guidelines for PKI and secure socket layer (SSL) encryption must be established. These guidelines should be compatible with existing B2B browser-based implementations that use HTTP/HTTPS.

The B2B gateway will also need a set of complementary services, such as the following:

- Receipt and routing—a public processing area that receives, authenticates, validates entitlement, archives, and routes inbound messages
- Package and delivery—a public processing area that packages, encrypts, validates entitlement, digitally signs, archives, and delivers outbound messages
- Process automation and application integration—a private processing area that provides for process automation and backend integration of inbound messages and outbound messages
- Infrastructure for non-repudiation database (NRdb), trading partner database (TPdb) and PKI
- Notification services for e-mail, pager, etc.
- XML/HTML scraping—ability to extract data from remote trading partner Web pages, in addition to or in lieu of data passed within RosettaNet messages
- Trading partner portal—a portal where trading partners can self-administer their RosettaNet processes and subscriptions
- Testing facilities—the ability for trading partners to test their RosettaNet messages against a test site; after self-testing, the B2B gateway would then promote the trading partner from "test" to "production" status, and thereby allow trading partners to control their production messaging processes
- Satellite capability—a "host" could provide its non-automated trading partners with a B2B satellite solution, thereby acting as a hub, developing and supporting a B2B application at its trading partner facilities

Figure 5 illustrates an integrated B2B gateway, complete with support for RosettaNet, other B2B initiatives, and SMTP, FTP, and VAN/EDI.

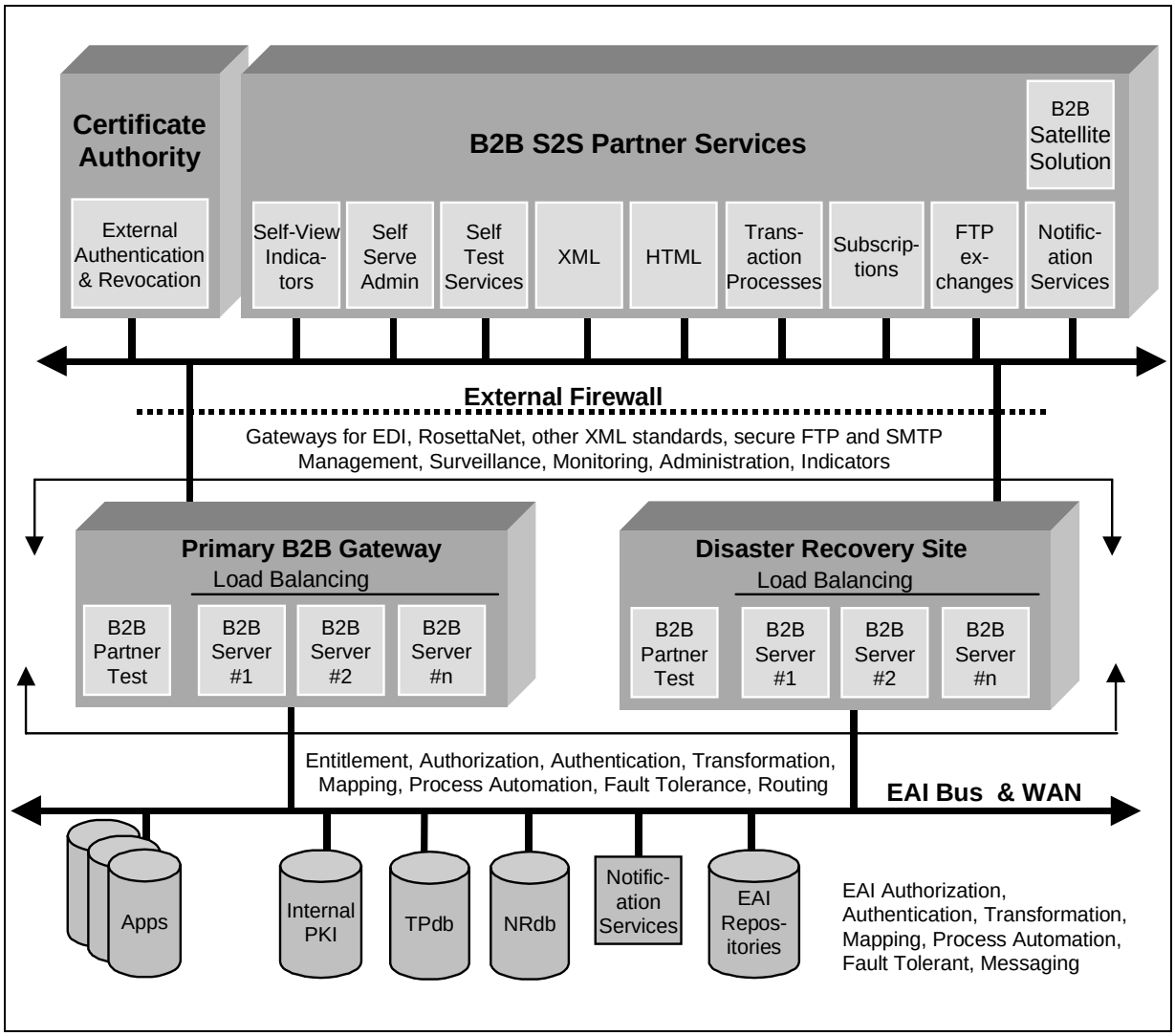



**Figure 5: An integrated B2B gateway**

Infrastructure readiness can be a huge task. However, it is not necessary to do it all at once. It is possible to install a solution from a RosettaNet solution provider within a few weeks and be up and running for a small implementation. Performing a comprehensive review of third-party solutions and then choosing a solution provider could take several months. However, even then one has only just started on the journey to overall infrastructure readiness.

### Business Process Readiness

The purpose of this track is to assess business and technical resources, legacy applications, and business process repositories. This is important because each RosettaNet PIP provides a mutually agreed supply-chain view of key business processes. Key criteria in this track include understanding the RosettaNet message structure and process message sequencing, identifying business process architects, and defining new business processes.

Intel is completing its initial B2B infrastructure planning and design requirements gathering, which coincides with initial RosettaNet pilots. A key finding from our initial efforts is that a significant increase in participation by "business analysts" (one of two areas in which we had difficulty obtaining resources) is needed in order to forward engineer and plan for the anticipated levels of adoption.

We have also realized that process re-engineering for RosettaNet must be performed within the context of re-engineering enterprise at-large business processes that

include requirements for trading partner integration. Therefore, our RosettaNet effort is considered an integral part of our EAI initiative. RosettaNet is therefore one element of a strategy to become a real-time event-driven global enterprise. Process re-engineering is a huge task.

To appreciate why business process readiness is such a big task, we need to understand how constructing a distributed Internet application using a robust message structure with a rich meta-model impacts enterprise readiness.

A RosettaNet message is intended to be predictable (open standards-based format), somewhat human readable, and portable between trading partners. In order to produce a widely supported and long-lived message format, the RosettaNet consortium agreed to define a message structure incorporating a complete data and meta-data model common to the significant business processes within the IT and EC supply chains.

A RosettaNet message consists of several nested XML structures and data structures, namely,

1. nested XML envelopes to define action, transaction, service, agent, message, transfer, and security sections
2. XML message sections for preamble, header, and body
3. attributes expressed using XML tags based on a supply chain dictionary
4. meta-data schema structures expressed using XML document type definitions (DTDs) or XML schema consisting of attribute data type definitions, tag hierarchy, cardinality (1:1, 1:n), permissible values, and parent/child dependencies
5. data as message content

Therefore, this message was deliberately designed as a self-contained, stateful and intelligent message, complete with data, persistent state information, and a meta-data model. Conceptually, it could be used to populate an object class or produce a database structure. Moreover, it could be abstractly considered as a snapshot of a transactional sequence in a file-based database expressed using XML.

It is therefore important to recognize that a RosettaNet message contains more information than data alone. It is a rich, fully stateful, self-describing package of information.

A RosettaNet message does not include any implied, hard-coded positional, or delimited structures. On the contrary, other formats for message and document exchange (namely EDI and non-standard comma-separated values (CSV) or tab-delimited file formats) provide a lesser degree or no level of schema definition, data constraints, dictionary-driven taxonomies, and process state information.

The completeness of a RosettaNet message structure across a supply chain (as defined in PIPs) requires significant forward engineering by trading partners within the RosettaNet consortium. As a result, trading partners should expect to re-engineer their back-end systems to become RosettaNet compliant. This may involve creating processes that currently don't exist internally or mapping processes that are currently different from RosettaNet processes.

Up-front business process architects need to participate in many activities:

1. RosettaNet PIP workshops to define each process, meta-model schema, dictionary, taxonomy, message sequencing, and run-time parameters (e.g., wait times, retry duration, acknowledgements)
2. determining impact upon existing business processes and existing applications
3. optimizing existing business processes by leveraging the capabilities provided by RosettaNet within the context of an at-large enterprise process re-engineering effort
4. determining new processes and data services

### Application Development Readiness

The purpose of this track is to prepare PIP implementation development plans and roadmaps. This is important because this step represents how and when existing processes and systems will be modified and rolled out to support RosettaNet. Key criteria in this track include statements of work, budgets, and plans.

As in the Business Process Readiness track, substantial participation by application development group(s) is necessary to forward engineer and plan for the anticipated levels of adoption. Application development groups realize they need to re-engineer processes for RosettaNet within the context of re-engineering enterprise at-large business processes while at the same time including requirements for trading partner integration.

Key deliverables for this track include work scope; identification of impacted systems; identification of key business analysts and process architects; determination of RosettaNet compatibility with existing processes; preparation of project budgets and schedules; setting of release dates; provision of consolidated test

requirements; definition of necessary API components; and setting of incremental upgrade roadmaps.

This track is similar to most enterprise application development efforts and can use a variety of development methodologies (e.g., traditional waterfall, rapid application development (RAD), etc). This track, more than any other, is likely to require the greatest amount of effort and resources. What's important to understand is that this group is usually the last to participate in the RosettaNet implementation planning effort, yet it has to be the first to implement the plan in order for deployment to progress. Therefore, getting up-front participation from the application developers is mandatory.

## B2B EXTERNAL INITIATIVE READINESS

The purpose of this track is to assess the completeness and usability of the work of the chosen B2B external initiative (in our case, RosettaNet). Specifications, policies, and architectures provided by the initiative must be understood and evaluated against internal policies, procedures, guidelines, and strategies. This is important because implementing RosettaNet is not "only" a technology; it is part of a strategy that must permeate an enterprise's trading partner integration strategy. Key criteria in this track include review of consortium supply chain, implementation framework, and process frameworks.

Each B2B initiative provides technical specifications that present the functional design and technical frameworks for message structure, message transport, and/or message content. In the case of RosettaNet, many technical documents and specifications have been written. For example, below is a collection of guidelines and specifications that are necessary in our implementation of the "Manage Purchase Order" PIP (which covers submit, acknowledge, change, and cancel purchase orders). This material addresses one of approximately 100 PIPs.

1. RosettaNet Implementation Framework v1.1
2. Manage Purchase Order Specification (3A4)
3. 3A4 Purchase Order Acceptance Message Guideline
4. 3A4 Purchase Order Acceptance Guideline DTD
5. 3A4 Purchase Order Cancellation Message Guideline
6. 3A4 Purchase Order Cancellation Guideline DTD
7. 3A4 Purchase Order Change Message Guideline
8. 3A4 Purchase Order Change Guideline DTD
9. 3A4 Purchase Order Request Message Guideline
10. 3A4 Purchase Order Request Guideline DTD
11. Preamble Part Message Guideline
12. Preamble Guideline DTD
13. Service Header Part Message Guideline
14. Service Header Guideline DTD
15. Acceptance Acknowledgement Message Guideline
16. Acceptance Acknowledgement Guideline DTD
17. Acceptance Acknowledgement Exception Message Guideline
18. Acceptance Acknowledgement Exception Guideline DTD
19. Receipt Acknowledgement Message Guideline
20. Receipt Acknowledgement Guideline DTD
21. Receipt Acknowledgement Exception Message Guideline
22. Receipt Acknowledgement Exception Guideline DTD
23. General Exception Guideline DTD
24. General Exception Message Guideline

A given consortium's documentation is usually targeted to a specific supply chain or e-Business market segment. The consortium's pervasiveness within its target markets must be considered. Moreover, due to the relative youth of Internet e-Business, frameworks and specifications may not be as complete or thorough as they could be. Therefore, participation in and achieving time-tested experience within the initiative enables trading partners to more accurately assess the applicability of the initiative to their businesses, as well as providing a means for influencing the initiative such that it *does* deliver the needed benefits. Finally, adopting a B2B framework needs to include a review of its compatibility with best known methods (BKMs) within one's company.

### Trading Partner Readiness

The purpose of this track is to assess the readiness of key trading partners. This is important because one cannot implement RosettaNet without at least one and hopefully many trading partners ready to do so. Key criteria in this track include selecting trading partners,

choosing processes, detailed integration, and achieving reliable results.

Each trading partner will need to provide a similar level of effort. It will be several years until the B2B trading partner automation technologies have matured to provide relatively inexpensive plug and play solutions; therefore, these next few years will only include trading partners who consider themselves early adopters. Trading partners must have the will and desire to deliberately re-engineer business processes based upon a rapid schedule and evolving processes. They must be able to move quickly, often with ad hoc funding and scavenging for equipment and resources. Although management commitment is essential to successful implementation of a RosettaNet-sized initiative, a skunk-works and entrepreneurial mentality in the early days can be helpful.

Selecting a RosettaNet trading partner is currently easy because only early adopters are playing; and, with a limited set of PIPs to choose from, it is easy to define a project. A key expectation is that the use of RosettaNet specifications will eliminate the currently high level of up-front trading partner analysis needed to conduct e-Business. This may lead to a rush of trading partners wishing to engage each other using RosettaNet processes (after initial successful implementations by early adopters) before the PIPs have matured and PIP implementation is a widely understood experience. At present, early adopter trading partners spend significant effort figuring out how to use the RosettaNet specifications with one another. Once sufficient infrastructure is in place, the full benefits of RosettaNet can be realized as trading partners self-administer their processes and subscriptions.

Currently, readiness must be planned with exact testing and production dates and known versions of specifications and guidelines. Legal issues need to be negotiated up front (see "Legal Readiness" below). Precise details of Global Trade Identification Number (GTIN), United Nations Standard Products and Services Classification (UN/SPSC), and Dun & Bradstreet-assigned unique corporate identifier (D-U-N-S*) must be managed. Personalized trade parameters such as part number, product lines, and interpretations of timeouts, retry and acknowledgements need to be exactly discussed. Trading partner agreements (TPAs) need to be signed. Current EDI processes with the trading partner may need to be changed. Digital certificates and digital signatures will be needed. And, as always in a new venture, backup plans will be needed.

### Solution Provider Readiness

The purpose of this track is to assess the readiness of your selected B2B gateway solution provider. This is important because the tool you have selected may not provide all the capabilities needed to implement a PIP with trading partners. Key criteria in this track include review of public and private PIP processes, review of PIP templates, and concurrence of PIP interpretation.

Some solution providers provide only the plumbing to enable RosettaNet. When no PIP templates are provided, the end user must provide all aspects of PIP implementation. In these cases the tool is ignorant of the exact meaning of retry periods, duplicate messages, acknowledgements, failure to receive, and other process specifics. These build-your-own solutions will require internal infrastructure for non-repudiation database (NRdb), trading partner database (TPdb) and PKI.

Other solution providers provide a robust framework for PIP implementation where the PIP template is quite cognizant of the PIP framework. PIP implementation would be easier and faster using these tools; however, the tool must be sufficiently flexible should the PIP framework prove incomplete in any given trading scenario. These all-encompassing solutions include infrastructure for NRdb, TPdb, and PKI.

Trading partners need to assess the capabilities of their solution provider(s). Some key questions include the following. What level of compliance does the tool provide for the implementation framework and process specifications? When is beta and general availability? Has the tool been sufficiently stress-tested for a variety of PIP scenarios? Does the tool provide diverse role-based control so different groups cannot access other groups' processes? Many other questions will be on the minds of individual trading partners.

Finally, RosettaNet is working on a Solution Provider Certification program and certification standards, which should help RosettaNet implementors perform their assessments more quickly and with greater assurance.

### Legal Readiness

The purpose of this track is to assess relevant legal issues. This is important because RosettaNet has expanded the capabilities of trading partner integration beyond the current terms and conditions found with EDI agreements; therefore, legal precedence has not yet been established for RosettaNet interactions. Key criteria in this track include trading partner agreements and early participation by legal counsel.

* Third-party brands and names are the property of their respective owners.

Performing RosettaNet message exchange with trading partners will require a trading partner agreement (TPA) between each pair of trading partners. These legal agreements need to be managed by each company's legal counsel. TPAs currently exist for EDI; however, a generalized RosettaNet TPA does not exist as of this writing (although creation of a model TPA is now underway). In addition, legal expertise for Internet-based e-Business using RosettaNet has not yet been attained. Experience gained in RosettaNet pilot programs will help legal counsel to understand the differences between RosettaNet and EDI and facilitate the preparation of a comprehensive TPA.

Because RosettaNet will be enabling supply chain automation across a lengthy chain of buyers (customers) and sellers (suppliers), the goal is to write the TPA from a neutral perspective. Use of such a neutral TPA may be a challenge for many companies, whose organizational practices may have dictated that they prescribe different terms and conditions within their EDI TPAs depending upon their role as buyer or seller.

The list of legal concerns is being compiled as we move forward. Although many issues have been identified, the full impact will likely not be comprehended until the infrastructure is in place and more time is spent in understanding legal ramifications. To date, some of these issues are

- encryption export to controlled countries
- frequently changing e-Commerce and e-Business legislation
- strict privacy laws
- the potential for hundreds of trading partners with varying capabilities
- restriction on use of confidential, proprietary, or trade secret information
- constantly changing landscape of trading partners, processes, messages, documents
- personalized TPAs with specific and different run-time parameters
- self-administered processes and subscriptions
- proper use of digital certificate and signatures for the accompanying document/message
- signed non-disclosure and confidentiality agreements

### Security Readiness

The purpose of this track is to assess the security requirements for encryption, authentication, and authorization at both the network and the trading partner message exchange level. This is important because implementing RosettaNet means that trading partner systems penetrate their corporate external firewall and security mechanisms. And undoubtedly, most data will need to pass through the internal firewalls to core enterprise applications. RosettaNet also will enable trading partners of different types and privileges to exchange documents for many critical business processes (e.g., purchase order, quotes, product information, pricing, availability, inventory, technical specifications, trade secret and confidential documents, CAD drawings, design specifications, etc.). Key criteria in this track include an understanding of corporate security and document confidentiality policies; and encryption, authentication, and authorization.

Security needs to initially address the front-end and the back-end. Front-end security issues apply to firewalls, proxy servers, network routing and protecting the system from malicious attacks. Back-end security issues apply to the controlled access to message content to system users and intermediaries using a right-to-see approach. Unlike current point-to-point solutions where data handling is decentralized, a B2B gateway will provide for a centralized flow of critical business information; therefore, only users with the right to see specific data should be entitled. Role-based administration of the B2B gateway should be considered.

Security readiness is also a significant challenge due to the inherent solution complexities, need for managed risk, and elevated concerns. The RosettaNet implementation framework incorporates a public key infrastructure (PKI). Intel's current RosettaNet implementation is based on a single corporate guideline using multiple certificate authorities, digital certificates, and digital signatures. Intel also requires the use of 128-bit encryption, which is greater than common usage and also is prohibited for export to controlled countries. Obtaining, understanding, and incorporating these guidelines and technologies into the B2B gateway, although logically simple, has been technically difficult due to the inherent complexity of PKI.

An important aspect of security is the ability to immediately revoke the privileges of a trading partner, or of any of their processes or subscriptions. It is also important to be able to confirm that trading partners are sending messages as agreed. This includes being able to detect when a message was not correctly assembled and

transported according to the TPA in place between the trading partners. This also includes the ability to detect whether encryption, digital certificates and digital signatures were correctly used.

### Audit Readiness

The purpose of this track is to assess one's readiness to be audited by internal company officials. This readiness is important because RosettaNet trading enables the interaction of critical business processes. Managers and executives should not be casual with their views of implementing RosettaNet. Key criteria in this track include understanding the seriousness of global electronic trading, and preparing for audits.

Knowing that a B2B gateway will eventually transport and manage a majority of e-Commerce transactions and e-Content interactions with trading partners, it is important to design the B2B infrastructure up front to withstand frequent and diverse auditing.

By design, RosettaNet and the capabilities it enables represent considerable risks to a company should something go wrong. Auditing is actually a good thing, as one should feel more assured that risks are under control. Some of the risks identified include

- potential to be majority revenue channel
- binding $M transactions
- binding legal agreements
- international trade with an easy global reach
- rapidly changing trade and Internet laws
- many government enforcement authorities
- sensitive and confidential document/information exchange
- many micro projects with intangible ROI where something will be unforeseen
- potential for lost potential or mistakes
- needs to be fault tolerant without data loss
- many critical success factors
- pivotal and timely information exchanges
- potential for significant impact on internal systems
- significant visibility and expectation levels
- competitors waiting for your misstep!
- centralized administration of enterprise processes and data (need for role-based administration and management using a limited right-to-see basis)

## CHALLENGES

Achieving a common language for e-Business offers challenges in a number of areas, including (but not limited to) the development of the specifications themselves; correctly identifying the internal barriers to success and successfully overcoming them; and planning to keep up with an ever-changing business, technical, and standards environment.

Some of the specific challenges we see ahead include

1. *Internet Speed*. RosettaNet* is caught up in the frenzy of Internet time. As such, trading partner automation and XML messaging are very hot technologies; the leaders in this race will likely reap the greatest rewards. Most significantly, getting it done faster, better, and cheaper will remain a requirement that cannot be understated. Many challenges exist when trying to compress and accelerate planning, funding, scheduling, evangelizing, designing, building, and testing, especially when considering the Readiness Model presented above.
2. *Sustaining will (internally).* Maintaining momentum in the face of "short attention spans" seems to be a systemic symptom of today's Internet e-Business mentality. At an increasing rate, everyone seems to have less time to make informed decisions. An increased level of risk-taking will be necessary to proceed; management needs to remain committed even when the inevitable mistakes are made.
3. *Sustaining will (externally).* Early adopters of RosettaNet will find it neither easy nor inexpensive to initially embrace. Each supply-chain or endorsing adopter will face an inevitable debate of whether to continue or disengage. So far the will of the RosettaNet consortium is withstanding these stresses and the key motives for moving forward remain steadfast; however, further tests of will are likely before RosettaNet's adoption is widespread.
4. *Obtaining resources*. Planning in advance for resource needs is a challenge within any company; however, RosettaNet, like all e-Business initiatives, is driven by its constituents faster than any company could anticipate. Obtaining business and technical resources is a challenge; however, expanding to include sufficient forward-thinking resources from business analysts, technical analysts, system architects, and application architects requires resource allocation. This can be achieved either by additional funding or by cancelling other planned projects. This can be especially challenging if resources are being pulled from competing B2B initiatives.

* Third-party brands and names are the property of their respective owners.

5. *Creating the implementation plan*. Defining, planning, and estimating the scope of work to implement which of the ~100 RosettaNet PIPs across the enterprise at-large requires a diverse group of resources and a PIP-centric approach rather than a business group approach.

6. *Choosing a project management methodology*. RosettaNet implementation needs to be executed using a hybrid of rapid application development (RAD) project methodology. Determining a methodology could be challenging within companies that do not have a conscious process for selecting a methodology.

7. *Finding an optimal team structure*. Initial implementations of RosettaNet require participation from diverse groups within an enterprise (exact composition depends heavily on the PIPs chosen for implementation). Each PIP implementation becomes a mini-project within the bigger context of RosettaNet and B2B implementation. Maximizing team productivity and effectiveness is essential, especially considering that B2B and e-Business projects need to proceed at Internet speed. It will be challenging to form an optimal team structure, then clone it for the many PIP mini-projects.

8. *Managing information overload*. Implementing any enterprise-wide project (especially one which happens to affect the very way the enterprise conducts its business) is hugely complex and involves a tremendous amount of information assimilation. Implementing the same set of specifications across most of the members of an industry magnifies the problem of synchronized information assimilation enormously. Participants in the implementation process must remain current with respect to RosettaNet specifications; each of the open standards on which RosettaNet is based (e.g., XML, SSL, HTTP); software and hardware solution options; internal company guidelines; requirements and functional specifications; test plans; meeting minutes; and other common materials. Participants must also keep abreast of similar materials from trading partners with whom they are implementing the plan. Methods for assimilating and managing frequent knowledge and information change in the e-Business sphere are sadly lacking.

## RESULTS

On a practical level, we have identified eight distinct roles within our B2B RosettaNet* deployment strategy. Table 1 lists these eight roles; it also shows the level of participation of each of these players within the readiness tracks discussed above. (As a point of departure for readers, the staffing levels for each role as we worked through to our 2.2.2000 deployment plans was as follows. One person each fulfilled roles 1 through 5. Role 6 consisted (in our case) of one full-time person plus parts of numerous other folks participating in PIP workshops, for another full-time equivalent. Multiple people participated for roles 7 and 8, typically one person for specific groups of PIPs or core applications. A total of 22 people participated for 2.2.2000 -- 13 from IT and 9 from the business units)

Intel performed several key tactical steps to address the diverse issues within the Readiness Model.

First we assembled the Intel RosettaNet Deployment Team consisting of six people in roles 1 through 6 in Table 1. We were slow in getting participants for roles 7 and 8 because these groups were extremely busy and up-front resource planning was required. In hindsight we recommend engaging these business analysts and application development groups in the early stages of RosettaNet planning.

Next, we engaged one trading partner (a major distributor) as part of our RosettaNet proof-of-concept pilot (August 1999) and initial implementation (2.2.2000). With our trading partner, we selected PIP3A4 ("Order Management"). Each of us selected our own solution provider and tools. This meant that four companies had to synchronize development and test plans. Since we were all first implementers, gaps and changes in the RosettaNet Implementation Framework and PIP guidelines needed to be ironed out. Infrastructure planning was a key focus from the beginning. Engineers within the core environment supporting EDI and our e-Business engineering groups worked together to integrate e-Business design requirements with existing EDI requirements. At present, we are creating a production environment that supports both EDI and RosettaNet running on Windows NT*.

* Third-party brands and names are the property of their respective owners.

* Other brands and names are the property of their respective owners.

After a few months of assessing RosettaNet readiness and formulating the Readiness Model, we prepared and sent a PIP assessment and business impact survey to all business groups having a need for trading partner automation. We are now waiting for enterprise-wide responses. These responses, and additional partner readiness discussions, will be reviewed and become the basis for our post-2.2.2000 rollout.

The following recommendations are provided to assist with first-time RosettaNet deployment:

- Look for a quick win: pick one strategic PIP with one partner. Plan for the process to take 2-4 months. Assign 4 to 6 people.
- Engage the solution providers, letting them educate you on B2B and partner integration architectures. Perhaps even contract with one of them to build a limited production pilot. Defer committing to your B2B vendor until a successful pilot is in production.
- Require the Business Manager and Technical Manager to hold weekly meetings to review progress and status.
- Incorporate the RosettaNet roadmap strategy within the company's overall B2B strategy.
- Include other B2B channels within the scope of the B2B gateway (e.g., secure file transfer, SMTP, EDI).
- Consider the impact on existing browser-based applications and partner portal strategies.

**Table 1 :Participation levels of key roles in readiness model tracks**

| Role # | Description | Readiness Track | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 Biz Strat | 2 Infra-structure | 3 Biz Process | 4 App Dev | 5 B2B Initiative | 6 Trading Partner | 7 Solution Provider | 8 Legal | 9 Security | 10 Audit |
| 1 | RosettaNet Business Program Management | **L** | M | **S** | S | M | **S** | M | M | M | M |
| 2 | RosettaNet Technical Program Management | M | L | S | S | S | S | L | L | L | L |
| 3 | PIP Management | S | M | S | S | S | S | M | M | M | M |
| 4 | Pilot Management | S | S | S | S | S | L | S | S | S | M |
| 5 | Application Integration Management | M | S | S | L | S | S | S | M | S | S |
| 6 | RosettaNet Standards-Development | M | M | S | M | L | M | M | M | S | N |
| 7 | Technical and Business Analysts, Business Process Analysts | S | M | L | S | S | S | M | M | S | S |
| 8 | Back-end Application Development Management | M | M | S | S | N | M | N | N | M | S |

Legend: L = Leader S= Significant Participation M = modest participation N = little to no participation

## CONCLUSION

Our team continually expands its understanding of what it takes to implement RosettaNet. As we complete a second-phase pilot, plan for future implementations, design the infrastructure, and expand our circle of influence, we foresee many new challenges. It is unclear when the rate of discovery of new issues and challenges will diminish. It is likely not to be until widespread trading partner/PIP implementation occurs in several years.

## ACKNOWLEDGMENTS

Special thanks are due to our fellow members of Intel's enterprise-wide RosettaNet Deployment Team and to the architects, strategists, and technologists working on e-Business for their frequent business, strategic, and technical contributions to our understanding of issues

related to the implementation of RosettaNet and B2B trading partner automation capabilities.

Intel RosettaNet Deployment Team Members: Colin Evans (SMG IM&E Director, eBusiness Architecture and Operations), Alan Court (SMG IM&E, RosettaNet Program Manager); Andy Keates (SMG IM&E RosettaNet Pilot (e-Concert) Manager); Thomas Vlach (PLG, RosettaNet PIP Analyst).

IT Architects, Strategists and Technologists: Bert Cave (IT Engineering, EAI Program Manager), Ralph Nitta (IT e-Business Integration Manager), Ed Balthasar (IT Strategy and Technology).

## REFERENCES

[1] RosettaNet specifications (www.rosettanet.org).

## AUTHORS' BIOGRAPHIES

Patricia J. O'Sullivan is a staff architect in IT Strategy & Technology. Her current focus is on increasing productivity in the use of information, currently as applied in e-Business standards. As part of this focus, she leads the RosettaNet's Implementation Framework team. She holds a masters degree in library and information science. Her e-mail is patricia.j.osullivan@intel.com

Don S. Whitecar, PE, is IT's e-Business EAI/B2B and RosettaNet Program Manager. His goal is to provide a robust, scaleable, and secure infrastructure for B2B trading partner automation. Don is also responsible for Intel's technical implementation of RosettaNet, working closely with business groups, business analysts, and application architects. His e-mail is: don.s.whitecar@intel.com.

# SHARED NETWORK GOVERNANCE AND STEWARDSHIP OF DATA AND THE EXCHANGE OF DATA

**A White Paper of the**
**INFORMATION INTEGRATION INITIATIVE**

Draft - June 21, 2000

---

## PURPOSE

This paper discusses the importance of shared governance and stewardship of integrated information. Shared governance and stewardship pertains to EPA, states, and other stakeholders in establishing and maintaining the environmental exchange network. It pertains to the program offices, EPA's Office of Environmental Information, and regional offices within EPA in establishing and maintaining the EPA node(s) on the network. It pertains to the data, the exchange of data, management of the databases (including registries), and maintenance of the network nodes of the national environmental information exchange network.

### Background:

Currently, each program office manages or coordinates the management of data that pertains to its programs. Except for some basic facility registration information, data are not integrated across programs. Data are either publicly accessible or not accessible at all. This situation has been driven by laws that focus on only one aspect of the environment at a time, separately delegated down to individual programs offices, who, given time and budget constraints, develop regulations and systems in support of the individual programs.

As citizen interest in local and global environments grow and as access to information becomes more an more available to citizens locally (mostly via the Internet) constituents want to have a broader, overall understanding of their environment and how it affects them. We are now being asked to integrate and harmonize data sets that up to now have been collected, maintained and disseminated in very much a stovepipe fashion. We will all need to work together to make this happen, and given time and budget constraints, there may be efficiencies to centralizing much of the activity under an agreed upon set of procedures, standards and protocols.

## DISCUSSION

### What is Governance and Stewardship?

Governance:

In an integrating data context - **Governance** is the development and implementation of a set of rules for managing the network including data standards, protocols for exchanging data, procedures for maintaining and improving data quality, agreements for preserving security including accessibility, integrity and confidentiality and particularly, agreements on who will maintain and make accessible the authoritative copy of each of the data sets.

Shared Governance:
Network partners need to participate in forming and ensuring compliance with the procedures, protocols, standards, etc. necessary to maintain each part of the network. It is important that the network partners come to consensus on the rules they will need to follow as network stewards. There will be governance of the network across network partners, as well as governance of the network node across the offices, regions, and other components within an organization that manages that node as a partner on the network.

It is not enough just to share the data, it needs to be properly maintained (quality assured, updated, accessible, explained, etc.). Good stewardship can be ensured through trading partner agreements, performance partnership grants, and performance partnership agreements as well as other measures. Where good stewardship can not be enforced, statements about data quality and availability can be made on the sites that point to the other sites on the network.

Stewardship:
In an integrating data context - **Stewardship** is managing the data, resources or activities – from data collection, through maintenance and disposition. It is foremost the role of quality assurance, but extends to the analysts' "respectful use of data" and includes making the data available to all those (and only those, in the case of proprietary and security related issues) who are authorized to access it. Stewardship extends, also, to efficient management and effective integration with other, related data.

Stewardship is not ownership. EPA, its program offices and regions, the states - none of us actually own the data. We manage the data for taxpayers, stockholders, etc.

Shared Stewardship:
Stewardship is a shared responsibility:

- Across all organizations in the network, each organization which provides data for use by others on the network must exercise stewardship over the data, but the roles and responsibilities will vary from mere warehousing to actually ensuring and maintaining the quality and timeliness of the data. However, without good stewardship, there may exist data of suspect quality, availability and usability which forms a gap in the web of data contained in the network. Mechanisms will need to be in place to coordinate and resolve problems with the flow and sharing of the data, costs vs. value added, data quality, etc.

- Within each organization on the network, stewardship is also a corporate or agency-wide responsibility. EPA, for instance, will need to work out stewardship of the following aspects of our network:

    - Data content, integrity, and quality - by data set, data table, field and record
    - Applications to collect the data
    - Applications to process the data
    - Applications to provide access to the data to the public and other stakeholders

- Maintenance of the database that houses the data
- Maintenance of the hardware and operating software

This can be difficult when the data were formerly "owned" by several of the offices within the organization. However, shared stewardship has economic and access benefits and can be accomplished through agreed upon protocols, processes and standards as well as access controls.

**Why do we need Shared Governance and Stewardship?**

Data collection and management is costly and is best shared: Data is, increasingly, a major environmental management resource. Because of its value, and because of its high cost, it must be preserved and used by all who want to, and have legal authority to, access it. In the words of EPA CIO, Ed Levine, "If I know you have certain data, and I know how to get access to it, and it's reliable, good quality data, then I can depend on your data without having to collect it or maintain it, too!"

Data quality improvements are to be shared: In addition, improvements and corrections to the data need to be shared so that everyone benefits and can rely on the shared data source.

Data needs to be where you expect it, when you expect it, of known quality, source, etc.: Using an agreed upon set of data standards, formats, metadata, etc. makes finding the appropriate data more efficient and using the data correctly more feasible. It allows consistent and reliable transfer of data using understood standards and via compatible mechanisms.

**What Kinds of Stewardship are Needed for I-3?**

Stewardship is a far-reaching concept. At a high level, the network as a whole requires stewardship both by the states and EPA. Each partner will be the steward organization for its own node on the network, making sure it is functioning properly and that the data are available through it within the jointly agreed upon terms.

Each organization that is exchanging data on the network is responsible for ensuring that its data are transmitted and received in the agreed upon format and timetable, that the integrity of the data are intact, and that, in the case of confidential or other sensitive data, the data have not been intercepted. Hence, there is a need for stewards in each organization.

Individuals within the partner organizations (e.g., EPA, each state) will need to be responsible for making sure the hardware on which the data reside, and the software that secures and serves up the data are all working properly. This includes operating software, database software, applications software, etc. Within EPA, there is a need to determine how to assign stewardship responsibilities. For I-3 purposes, we may find that stewardship of the data (and tables within the databases) as well as applications to collect data should be decentralized, while stewardship of the database engines, hardware and operating software as well as applications to access the data publicly or across the agency should be centralized.

There is also the concept of a custodian who merely warehouses a copy of the data for convenience of access without any effort to improve the quality of the data or participate in governance.

Registries (See the white paper "The Proposed Use of Registries in Information Integration" for details) on the network must have shared (corporate) stewardship across the relevant constituencies if they are to be reliable and authoritative sources for commonly used facility, corporate, industrial sector, place, chemical, etc. data.[1]

Programmatic or State Data Linked to (but not actually in) Registries:
Registries will link to data sets that are not actually part of the registries. It is essential that the links (e.g., EPA facility ID or program system facility ID - whatever is the agreed upon number, EPA chemical ID or CAS number - whatever is the agreed upon number) remain intact. However, the quality of the other data within their program or state records is entirely under their control. Stewardship for that data and how any data set on the network should communicate the quality and timeliness of the data in that data set (e.g., meta data about the data set or extra fields indicating when the record was last updated, etc.) would need to be defined in an exchange format and agreed upon in a trading partner agreement. This is especially important for secondary users of their data and that meta data need to be prepared and shared to support that type of use.

Transition to stewardship of registry data:
The first registry to be established will be the regulated facility registry (FRS), maintained centrally by EPA with the support of the stewardship network. Over time, EPA (and perhaps other entities) will add additional registries for place, chemical and other substances, and expand the facility registry to include other entities that are not part of the facility registry (e.g., corporations that do not have facilities in the facility registry). As each registry is added, it should become the authoritative source of that information and be integrated with the other registries and network databases as appropriate. For more information, see the "The Proposed Use of Registries in Information Integration" white paper.

Transition to stewardship of non-registry data:
Ideally, the vision is to move toward a network where each partner on the network maintains its own data on its own server in a database that is compatible with the rest of the network and its applications. Thus, except for the registries, states and program offices could maintain their own data locally and outside users of the network could access the information through applications that would reach across the network. However, because not all network stewards are ready to provide this sort of access in a

---

[1]The roles for registry data stewardship described in the *Facility Registry System: Data Steward Manual - May 12, 2000 Draft*, pages 4 through 8, could easily be adapted to cover other registries that have been proposed in the "The Proposed Use of Registries in Information Integration" white paper (mostly by changing "FRS" to "EPA registries" and "FLA" to "EPA registry tools"). The roles include a Data Stewardship Manager (for the Agency as a whole), EPA Program Data Stewardship Managers, Regional Data Stewardship Coordinators, Regional Data Stewards, and Participating State Data Stewards.

secure manner, there will clearly need to be a transition period where EPA centrally maintains accessible copies of non-registry programmatic and state data sets and acts as a custodian (without changing the data) to simplify and speed up access. For example, in the short term we may use data warehousing, whereas in the longer term we would move toward a decentralized access ("come and get it") approach, or more likely, some combination of the two over the exchange network.

**CONCLUSION/RECOMMENDATION:**

Stewardship and governance are shared responsibilities. All of us involved in the environmental exchange network will need to develop and implement data standards and rules of operation that support the overall goals of the network, or it will falter.

Some data sets may originate from numerous sources, but in order for them to be fully integrated, will need to be managed centrally based on mutually agreed upon data and procedural standards. OEI within EPA may take on that role for certain data sets, but other program offices, certain states or other stakeholders may be the appropriate organizations to take on that role for other data sets.

Although each organization using the network will need to sustain a certain level of trust about the other parts, verification methods may be necessary to ensure that all data sets on the network continue to be maintain an acceptable level of quality, availability and security and metadata.

A successful I-3 network will require an associated stewardship network. We recommend that the I-3 stewardship network be expanded from the current Facility Registry System (FRS) effort to cover other registries, and that non-registry data stewarded by States and program offices be included in that stewardship network, although to a lesser extent.

## APPENDIX:

Data vs. Database vs. Application vs. Network Node Stewardship:
In addition to data stewardship is the concept that there is hardware and software that serve up that data to the network. The data stewards may do a fantastic job of cleaning and maintaining the data, but if the Internet server is down, it will be inaccessible. Also, the data may be in great shape, and the server may be working just fine, but the applications that provide access to the data are not user friendly or incompatible with other applications on the network.

What are the right formats in which the data should be stored for access? Most of EPA's data are stored in Oracle databases, but XML appears to be the format for exchange of data over the Web. However, Oracle uses relational tables, whereas XML uses a hierarchical structure. Also, the format for geographic data may be different from that format for other data sets. Clearly, database standards as well as data standards are part of the solution.

In addition, there will be various types of users on the network, some of which will have access to all the data, and some of which will only have access to public data. The **database steward** will need security controls at the database level to control access by various types of users.

In order for the database security to work correctly, there will also need to be appropriate security at the server level. A **network node steward** will need to make sure these security measures are in place as well as making sure the server is up and running and connected to the Internet.

Once the data is in (a) standard format(s) on Internet servers with the appropriate security controls, there will need to be applications that give users access to that data. To the extent that the access applications are on the network node that provides the data, there will need to be an **application steward** to keep that application up and running, up-to-date, etc. and compatible with the rest of the network. To the extent that the access applications are on the user side, they are the responsibility of the user and there would not need to be an application steward.

Example: Facility Registry System (FRS):

Facility data are collected for the following programs: TRI, RMP, RCRIS and BRS (RCRAInfo), CERCLA, PCS, SDWIS, AIRS, NCDB.
In many cases the collections come through the regions; and in many cases, the states.

FRS data stewards need to include representation from each of those program offices as well as regions and states. Note that program offices that do not collect that type of data do not have obligations for that registry (e.g., OPP).

For the data stewardship to work, users on the network need to:

- Know where to find what data - so there needs to be a place from which to access all registries that is widely advertised,
- Be able to reliably access the data - so there need to be set hours of operation (7 days a week, 24 hours a day or just regular business hours?), as well as servers that can handle the level of traffic expected, and clarity on what portion of a data stewards data each user has access to (states may be able to access more than a member of the general public)
- Know the meaning, quality and currency of the data - so there need to be data standards and meta data that include definitions of each data element, interpretive data, measures of data quality that are in turn explained, and information on when the record(s) was(were) last updated, as well as procedures for populating, updating, and accessing data sets.